A comparative eye-tracking study in great apes and humans: the pattern of eye movements for scenes and faces

Thesis by

Fumihiro Kano

Primate Research Institute, Kyoto University

November, 2011

Contents

Chapter 1: General Overview

A comparative eye-tracking study in great apes and humans: the pattern of eye movements for scenes and faces

1.1. Introduction

1.1.1. Technical eye tracking issues in apes

Despite the long history of eye movement research in humans and macaque monkeys, eye movements in nonhuman apes had not been studied until recently. This was largely due to the lack of appropriate eye tracking methods for apes (eye tracking: direct measurement of eye movements). The traditional eye tracking method is intrusive or even invasive, including wearing contact lenses (typically for humans) or implanting magnetic search coils (typically for nonhuman animals) on the eyes. In addition, most eye tracking methods require that the subject's head is restrained using a chin rest or a bite bar. These methods are not suitable for modern ape laboratory studies, in which the experimenters avoid invasive methods and establish a relaxed relationship with apes. However, due to recent advances in video-based eye tracking techniques, several types of eye-tracker have allowed us to record subject's eye movements remotely (i.e., *via* an infrared video camera) without its head restrained. We have applied this technique to most of ape subjects in a typical laboratory setting relatively easily, even in a setting in which apes and human experimenters are separated by transparent acrylic panels for safety purposes. In fact, this eye tracking technique has been applied to laboratory settings of at least three different ape facilities (Hirata, Fuwa, Sugama, Kusunoki, & Fujita, 2010; Kano, Hirata, Call, & Tomonaga, 2011; Kano & Tomonaga, 2009).

1.1.2. Reasons to study eye movements in apes

There are at least two reasons to study ape looking behavior. First, as the pattern of eye movements sensitively reflects how visual information is retrieved from the outside world and processed, comparing overall pattern of eye movements between human and nonhuman primates provides an insight into the evolution of strategies how they process visual information. In particular, apes are the most closely related to humans; thus, the direct comparison of eye movement characteristics between apes and humans would highlight unique aspects in humans.

Second, as the direction of gaze (or of the eyes) plays an important role in primate social communication, comparing social looking behavior between human and nonhuman primates would be a great value in examining the evolution of social communicative strategies. The primary function of gaze is to retrieve information. Thus, orienting gaze to significant social stimuli, such as faces, is beneficial to retrieve information about social partners, such as identity, sex, age, and emotion. In addition, gaze is informative communicative tools among social partners; the direction of gaze reveals the direction of interest to social partners regardless of intentionally or unintentionally. For example, eye contact, i.e., looking into a social partner's eyes, signifies a variety of social intentions, such as threat, sexual attraction, play, and reconciliation in primates (Argyle & Cook, 1976; De Waal, 1990; Goodall, 1986; Thomsen, 1974). Gaze following, i.e., tracing the gaze line of a social partner, plays an important role in complex social skills, such as cooperation and theory of mind, in primates (Baron-Cohen, 1995; Call & Tomasello, 2008; Emery, 2000; Tomasello, Hare, Lehmann, & Call, 2007). Apes have been well studied with regard to these topics and are known that they have highly sophisticated social skills comparable to those of humans. However, as it is practically difficult to directly observe the subtle looking behavior (i.e. eye movements), the detailed information about social looking behavior (e.g. which part of face is viewed; when and how long a certain part of body/face is viewed) is unknown; thus, the eye tracking method is necessary.

1.1.3. Scope of this overview

The aim of this overview is to show (1) how noninvasive eye tracking method can be applied to apes, (2) how apes and humans are similar and different in their eye movement characteristics, and (3) how comparative analysis of eye movements can contribute to an understanding of the evolution of visual information processing and social communicative strategies in hominids. Many previous eye movement studies on nonhuman primates (typically macaques) aimed to understand mechanics and neural circuitries controlling eye movements (Munoz & Everling, 2004; Robinson, 1968).

Nonhuman primates are also used in psychological studies on fundamental functions of eye movements; how the saccade (rapid jerk-like movements of the eyes that shift the gaze from one location to another) is generated in response to the presentation of simple stimuli (Findlay & Walker, 1999; Fischer & Weber, 1993). Both spatial (saccade direction and amplitude) and temporal (saccade latency) aspects of eye movements were measured. Typically, the subjects were extensively trained in the previous studies to make a saccade to (or against) the presented stimuli. In general, striking similarities have been shown between human and nonhuman primates in these fundamental functions of eye movements. Several studies have examined the free viewing patterns for naturalistic complex stimuli in nonhuman primates (Berg, Boehnke, Marino, Munoz, & Itti, 2009; Keating & Keating, 1982; Nahm, Perret, Amaral, & Albright, 1997; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010). Striking similarities were also shown between human and nonhuman primates in these free viewing patterns. However, several differences were also suggested, including general patterns of eye movements and scanning patterns for social stimuli such as faces (Berg, et al., 2009; Shepherd, et al., 2010). To reveal the phylogeny of eye movement characteristics among primates, we need to obtain more comparative data, especially those between humans and the most closely related primates, great apes.

Eye movements are known to be controlled in two levels; low-level and high-level. Low-level control of eye movements includes the reflexive or automatic saccades generated by the peripheral events (e.g. sudden appearance of stimuli) and the gaze drawn to the visually salient stimuli (e.g., white spot on a black background). On the other hand, high-level control of eye movements includes how flexibly the onset of saccade is regulated and how systematically the gaze is guided to the contents of stimuli (e.g., socially significant stimuli) rather than the visual properties of stimuli *per se* (Findlay & Walker, 1999; Henderson & Hollingworth, 1999). In this study, we were especially interested in high-level rather than low-level control of eye movements in humans and

apes.

The following section describes a method for recording eye movements of apes and humans. The third section provides specific results from our studies. In particular, we first examined overall similarities and differences in eye movement patterns for naturalistic social pictures between humans and the most closely related apes, chimpanzees. We then examined these similarities and differences more specifically in terms of low/high-level influences on eye movement patterns and also in terms of gaze responses to social stimuli, such as faces, facial expressions, and gaze cues of conspecifics (own species) and allospecifics (other species). We then examined the pattern of eye movements in other apes, gorillas and orangutans, using the same methods to compare the results with those of chimpanzees and humans. In the final section of this overview, we discuss how the results can be related to strategies of visual information processing and social communication in apes and humans. We believe that the studies on ape eye movements would pave a way to reveal ape cognition and would provide further evolutionary perspective for vision, visual attention, and eye movements.

1.2. General method

1.2.1. Noninvasive eye tracking technique with subject's head unrestrained

The key issue in tracking the eye movements of apes is to measure eye movements noninvasively with their head unrestrained. As shown in Figure 1-1, the experimenter and ape subjects are separated by transparent acrylic panels and ape subjects are not under the direct control of the experimenter. Therefore, ape subjects may take various viewing postures, move their heads during recording, or may stop viewing the stimuli in the middle of the presentation. Tracking their eye movements constantly in such an environment was challenging. To solve these issues, three factors were important: A video-based eye-tracker with a wide-angle lens, a flexible and interactive experimental setup, and stimuli to provoke spontaneous interest by the ape.

Figure 1-1

We employed a recent noninvasive, video-based, corneal reflection eye tracking technique. In this technique, the eye-tracker projects an infrared light onto the subject's eyes, and the infrared eye-tracker camera records two reflections from the subject's eyes (Figure 1-2)(Duchowski, 2007). One reflection is from the surface of the eyes (i.e., cornea), which is relatively independent of the head movements of subjects. The other reflection is from the retina (i.e., the so-called "red-eye effect"), which directly indicates the direction of the eyes (i.e., pupil center). By calculating the vector difference between the centroids of two reflections, the eye-tracker estimates the gaze direction that is relatively independent of head movements. We used a recent eye-tracker product (Tobii X60; Tobii Technology AB, Stockholm, Sweden), which has a wide-angle lens, records both

eyes, and allows relatively large head movements while preserving relatively high eye tracking performance in terms of speed (60 Hz; twice as fast as a typical camera) and accuracy (< 0.5° error in a standard recording setting).

Figure 1-2

To increase the eye-tracking accuracy, it is necessary to keep the eye camera at a certain distance from the subject's eyes (typically 60 – 70 cm) and coaxial to the eye direction of subjects. We employed a flexible and interactive experimental setup to accomplish this. That is, first, the experimenter encouraged the ape subjects to approach the acrylic panel as close as possible (using fruit pieces). Second, the experimenter set the eye-tracker and the monitor (LCD, 17-inch) on a movable table and adjusted the distance and horizontal position and vertical angle of the eye-tracker with respect to the subject's head. Although most of the ape subjects stayed still once they sat in front of the acrylic panels (without particular training), some of them, particularly juvenile subjects, were so active that they often changed their postures and head positions. In such cases, we employed a nozzle and tube attached to the acrylic panel, which produced a drop of fruit juice, and thereby kept the head of the subject close to the panel while drinking the fruit juice.

To keep the ape subjects' interest to the stimuli, we employed complex naturalistic pictures rather than simple geometric patterns. In addition, we limited the number of trials conducted during a daily session (typically six trials). Although, thanks to these efforts, the most of the subjects indeed kept their interest to the stimuli, some (10-15% of all subjects, particularly adults) hardly viewed the screen and turned their heads soon after the onset of stimulus presentation. Such individuals were

excluded from our studies.

1.2.1. Calibration

Calibration is a process in which the acquired gaze data is associated with the actual points on the screen where the subject was looking. First, several reference points were presented sequentially at different locations on the screen to guide the subject to look at those points. The acquired gaze data (vector difference between two centroids of retinal and corneal reflections; see above) was then associated with the actual coordinate axis on the screen. For human subjects, an instruction was given to view the small dots that appeared on the screen, whereas a short video clip or a piece of fruit was presented to attract the ape's gaze. The calibration accuracy was inspected after completing the calibration process, and the calibration process was repeated if necessary. The accuracy check was conducted before every session and occasionally during the session (between trials). With this calibration method, we confirmed that typical accuracy (the average distance between recorded and intended gaze position) was less than 1° of visual angle in all species under standard recording conditions.

1.2.2. Comparison among species

As the aim of this study was to directly compare spontaneous eye movement patterns for scenes and faces between ape and human subjects, we tried to minimize the methodological differences between the species. First, both ape and human subjects were tested using the same apparatus. Second, the same sets of stimuli were presented. Third, we prepared experimental conditions in which the subjects viewed the stimuli as naturally as they could. They were not forced to view the screen with their heads unrestrained. No explicit instructions or training were given to human subjects except to view the stimuli freely. Although rewards were occasionally used for ape subjects to keep them close to the acrylic panel, no rewards were used to reinforce any of their scanning patterns for stimuli *per se* (apes obtained the rewards regardless of whether they viewed the screen

or not). The daily session was < 20 min in both apes and humans. Thus, while we conducted the whole session in 1 or 2 days for human subjects, we separated the whole session into several days for ape subjects (typically around 10 days, six trials/ day, because the preparation of recording, e.g., encouraging the subject to come close to the acrylic panel, usually took a longer time in apes than human subjects).

1.3. Patterns of eye movement for scenes and faces

1.3.1. Initial observations

We first examined the eye movements of chimpanzees and humans when they were viewing naturalistic pictures, including both social and nonsocial stimuli (whole-body figure of a chimpanzee or human and various background objects; Figure 1-3) (Kano & Tomonaga, 2009). We tested six chimpanzees housed at Primate Research Institute, Japan, and 21 human subjects (graduate students). We found both similarities and differences between chimpanzees and humans for both overall eye movement patterns and responses to social stimuli. Both species concentrated fixation on the body, particularly the face of a figure in the scene (Figure 1-4). In addition, both species fixated on the faces at first sight (immediately after picture presentation), consistent with previous studies in humans (Fletcher-Watson, Findlay, Leekam, & Benson, 2008; Hershler & Hochstein, 2005). In contrast, the two species differed in that the duration of each fixation (fixation duration) was shorter in chimpanzees than that in humans on average (Figure 1-3); chimpanzees generally scanned the scene more rapidly and more widely than humans (Figure 1-4). The fixation duration was longer in humans than that in chimpanzees, particularly when viewing faces (Figure 1-4). These results raised several possibilities. (1) Both chimpanzees and humans concentrate fixations on informative areas (such as faces) rather than on visually salient areas (in terms of low-level image properties). (2) Each species has characteristic pattern in the timing of gaze shift (or speed of scanning). (3) Each species has characteristic pattern in their scanning patterns for the social cues.

Figure 1-3

Figure 1-4

1.3.2. Fixation location: low-level saliency and informativeness

We then examined the effects of low-level image properties (e.g., color, contrast, orientation) on eye movement patterns (Kano & Tomonaga, 2011a). We manipulated the global characteristics of the same pictures as those used in the initial study. Manipulations included eliminating the color, scrambling the configuration, filtering the image frequency, inverting the orientation, and reducing scene complexity. The same six chimpanzees and 16 human subjects were tested with these manipulated pictures. Both species preserved a strong tendency for fixating on faces for them. Interestingly, even when the picture with the head removed was presented, both species fixated on the area where the face would have been. These results suggest that the gaze was guided by informativeness rather than low-level saliency in both species when viewing a scene including significant social stimuli such as a face.

We used a well-established computational saliency model to examine whether chimpanzees and humans differ quantitatively in their responses to low-level saliency (Itti & Koch, 2000; Walther & Koch, 2006). This model calculates local saliency for each of the low-level properties of an image, such as color, contrast, and orientation, and combines them onto a single "saliency map." Figure 1-5 shows an example of a saliency map generated from a naturalistic scene. Dark areas indicate high saliency values. We found that the saliency model did not fully predict the actual fixation patterns of each species, suggesting that the subjects' gaze was indeed guided by informativeness rather than low-level saliency. We then calculated the saliency values of each species' fixated area (i.e., the average saliency values around each fixation). We found no species differences in this regard; i.e., both species were quantitatively similar in their sensitivity to local

saliency of the image.

Figure 1-5

1.3.3. Timing of gaze shift: fixation duration, saccade latency, and object tracking

We then examined the timing of gaze shift in chimpanzees and humans (Kano, et al., 2011; Kano & Tomonaga, 2011b). The same six chimpanzees and 18 humans were tested. We presented to them the naturalistic pictures including various image content (animal, fruit, object, texture) (Figure 1-6a). We found that both species showed a frequency distribution of fixation duration that was a skewed bell shape with a peak around 200 – 300 ms and a long right tail > 300 ms (Figure 1-6b); i.e., the fixations were typically for 200 – 300 ms but were often > 300 ms in both species. The same pattern of distribution has been repeatedly reported in previous studies in humans (Henderson, 2003; Henderson, Weeks Jr, & Hollingworth, 1999; Rayner, 1998). A clear species difference was observed; the distribution was skewed more to the left in chimpanzees than in humans because fixation > 300 ms was more frequent in humans than in chimpanzees (i.e., longer right tail in humans). Thus, on average, fixation duration was shorter in chimpanzees (about 230 ms) than in humans (about 330 ms), and the saccade was more frequent in chimpanzees (about four times per second) than in humans (about three times per second). This species difference was relatively independent of the image content (Figure 1-6a).

Figure 1-6

The subsequent question involves whether this species difference reflects their specific behavioral strategies or low-level automatic differences (e.g., eye muscles, retinal sensitivity). To examine this issue, we used a version of the gap-overlap paradigm (Figure 1-7a) (Hood & Atkinson, 1993; Landry & Bryson, 2004). In this paradigm, a central (fixated) and peripheral target stimulus appear in this order on the screen, and the subject freely views the stimuli. The time between onset of the peripheral target and initiation of a saccade to the target was measured (i.e., saccade latency). The subjects were tested under two conditions, gap and overlap. Under the gap condition, the central fixation stimulus disappeared shortly before the peripheral target appeared, whereas under the overlap condition the central fixation stimulus remained. To determine the effect of image content, we used either a face or an object as a stimulus in each trial. Consistently with the previous studies, humans showed the shorter saccade latency under the gap than overlap condition (i.e. gap effect; Figure 1-7b). However, this effect was weaker in chimpanzees than humans. As a result, the species difference in saccade latency appeared under the overlap but not gap condition; both species showed the same symmetric bell-shaped distribution for saccade latency, with a peak at 200 – 300 ms, under the gap condition (Figure 1-7b). That is, the species difference in saccade latency was observed under the condition when the gaze was detached from the central stimuli by the subjects themselves (overlap condition) but not under the condition when the gaze was forcibly or automatically detached from the central stimuli by the disappearance of them (gap condition). Consistent with the abovementioned results, we found no significant effect of image content on the species difference.

Figure 1-7

As shown above, chimpanzees scanned the scenes more quickly and widely than did humans. This may indicate that chimpanzees viewed more of the scene objects than did humans. To examine this issue, we modified the gap-overlap paradigm so that multiple small images appeared/disappeared at random locations on the screen at the relatively fast presentation rate (Figure 1-8). The subjects freely viewed the screen. To see the effect of image content, we used either a face or an object as the stimulus in each trial. As expected, we found that chimpanzees tracked the images better than humans; chimpanzees fixated on a larger number of images and viewed those images at the point of fixation for a longer time than humans. Consistent with the abovementioned results, we found no significant effect of image content on the species difference.

Figure 1-8

1.3.4. Responses to conspecific and allospecific faces (and facial expressions)

We then examined how chimpanzees and humans view each facial feature (e.g., eyes, nose, mouth, and ears). The same six chimpanzees and 18 human subjects were tested. We presented to them the close-up pictures of chimpanzee and human faces (Kano & Tomonaga, 2010). Figure 1-9 shows their typical facial scanning patterns. Both species viewed the main features (i.e. eyes, nose and mouth) rather than the other parts. Moreover, both species typically viewed the eyes first (immediately after picture presentation). A clear species difference appeared in the scanning patterns of eyes; humans consistently showed alternate scanning of left and right eyes, whereas chimpanzees

lacked such a pattern. Instead, chimpanzees fixated on the mouth more frequently than humans. Thus, the typical scanning path of chimpanzees appeared more vertical than that of humans; humans showed an inverted triangular shape in their face scanning path (connecting both eyes and mouth), consistent with previous studies in humans (Walker-Smith, Gale, & Findlay, 1977; Yarbus, 1967b). These species similarities and differences were consistent across conspecific and allospecific faces. We also found that humans showed a strong viewing tendency for eyes even when presented with faces including mouth expressions (Figure 1-10). In contrast, chimpanzees altered their viewing patterns for faces depending on the expression. Thus, overall, both species showed highly systematic but characteristic face scanning patterns; human differed from chimpanzees in their prolonged and consistent viewing of the eyes.

Figure 1-9

Figure 1-10

1.3.5. Responses to conspecific and allospecific gaze cues

We then examined the eye movement patterns for the gaze cues of conspecific and allospecific faces in chimpanzees and humans (Hattori, Kano, & Tomonaga, 2010). Eight chimpanzees at PRI and eight humans (graduate students) were tested. We presented to them the pictures in which the chimpanzee/human models directed their head to one of two target objects (Figure 1-11). We found that both chimpanzees and humans followed the head direction of models,

but this tendency was stronger in humans than in chimpanzees. In addition, humans followed the head direction of both chimpanzee and human models, whereas chimpanzees only followed that of the chimpanzee model. These differences in gaze following may be related to the face viewing duration in each species. That is, in general, humans viewed faces (of both conspecifics and allospecifics) for longer periods than chimpanzees; compared to humans, chimpanzees viewed the faces only briefly, but they viewed faces of conspecifics for relatively longer periods than those of allospecifics. Thus, overall, prolonged viewing of faces may have enhanced the stronger gaze following tendency in chimpanzees and humans.

Figure 1-11

1.3.6. Eye movements of other ape species: gorillas and orangutans

To examine whether the abovementioned similarities and differences between humans and chimpanzees could be generalized to those between humans and the other more distantly related great apes, we tested saccade timing and face scanning in gorillas and orangutans housed at Wolfgang Köhler Primate Research Center, Germany, using the same experimental procedures. First, four gorillas, seven orangutans, and 16 human subjects were tested with the same gap-overlap paradigm (Kano, et al., 2011). Consistent with the above results, species differences were found under the overlap but not gap condition. Apes shifted their gaze with an earlier timing than did humans, similarly to chimpanzees.

Second, we presented conspecific and allospecific facial pictures to five gorillas, 10 orangutans, and 12 human subjects (Kano, Call, & Tomonaga, under review). Consistent with the above results,

all species viewed the main features rather than the other parts. In addition, all species viewed the eyes at first sight (immediately after picture presentation). Again, there were species differences between humans and apes in viewing patterns for eyes; humans consistently showed alternate scanning of left and right eyes, whereas gorillas and orangutans lacked such a pattern, similarly to chimpanzees.

1.4. General discussion

1.4.1. Homologous nature of eye movements in apes and humans

The eye movement patterns for scenes and faces were strikingly similar, and thus essentially homologous between apes and humans. Apes and humans were similar in both spatial and temporal aspects of eye movements; i.e., in both spatial distribution of fixation location and frequency distribution of fixation duration (Figure 1-4 and 1-6). Apes and humans did not differ even quantitatively in their sensitivities to low-level visual saliency (Figure 1-5) or in their timing of reflexive or automatic saccade (Figure 1-7). Apes and humans were also similar in their responses to social stimuli. They were highly sensitive to bodies and faces, and the main features of faces (eyes, nose, and mouth) and gaze direction of other individuals (Figure 1-9 – 1-11).

1.4.2. Explaining species differences

We found two notable species differences. First, fixation duration (or saccade timing) was shorter and more stereotyped in apes, whereas it was longer and more variable in humans. Notably, this species difference was not attributable to the low-level automatic differences between the species, because both showed the same saccade timing (symmetric bell-shaped distribution) under the gap condition, i.e., when their gaze was forcibly detached from central stimuli *via* the disappearance of them (Figure 1-7). Thus, this species difference rather reflected the specific behavioral strategy in each species. Object tracking performance, as described in Figure 1-8, suggested that these strategies are in a trade-off relationship. That is, apes may employ more stereotyped but shorter fixation duration thereby enabling quicker scene scanning, whereas humans may employ a longer but more variable fixation duration thereby enabling flexibility in foveal analysis. In other words, apes and humans may differ in their behavioral strategies for how they allocate the limited time for the foveal analysis. This trade-off in the eye movement strategies may be related to the trade-off in the information processing; i.e. depth (human) vs. breadth (ape).

Alternatively, apes may have a poorer understanding of the contents of visual stimuli than humans, and thus were less engaged in each stimulus so they more quickly shifted their gaze from each stimulus than humans. However, this is unlikely because we found that content of visual stimuli did not significantly affect the species differences in overall fixation duration pattern. Note that the same species differences appeared even when meaningless scenes were presented (Figure 1-6). Thus, although the duration of individual fixation was influenced by the content of fixated stimuli (e.g., longer fixation duration for faces than background objects), the variability of fixation duration *per se* may have been determined by an internal mechanism rather than by the external stimuli.

Second, human face scanning was characterized by prolonged and consistent viewing of the eyes compared to that of apes. This tendency of humans was so strong that they did not change this tendency even when presented with faces with mouth expressions (Figure 1-10). One possible explanation for the distinctive human pattern of viewing the eye region is that eyes may have evolved additional communicative functions not found in other ape species. Compared with other primates, human eyes have a notable black-white contrast between the iris and sclera (Kobayashi & Kohshima, 2001). Moreover, fine motor control of the muscles around the eyes (Ekman & Friesen, 1978) enables humans to communicate a variety of directional and emotional cues. Thus, active viewing of the eyes may benefit humans by efficiently retrieving such information. That is, humans may use the eyes for bidirectional communication between senders and receivers. Prolonged eye viewing in humans may also enhance the tendency to follow gaze, as shown in Figure 1-11. In support of this suggestion, a previous study showed that apes are less sensitive than human infants to the gaze directions of a human experimenter when only eye directional cues are provided (i.e., when the head was immobile). An alternative interpretation of the active viewing of the eye region in humans is that eyes are more important than other facial features during facial processing; thus, intense fixation on the eyes by humans reflects their superior ability to process faces. However, the

previous studies do not support this interpretation; those studies showed that identity recognition is related to intense and systematic fixation on the entire complex of inner facial features, and not necessarily the eyes among humans (Hsiao & Cottrell, 2008; Malcolm, Lanyon, Fugard, & Barton, 2008; Stacey, Walker, & Underwood, 2005).

1.4.3. Eye movements of human infants, autistic individuals, and macaques

Several studies have examined the timing of gaze shift and the pattern of face scanning in human infants, autistic individuals, and macaques using a similar paradigm. The gap-overlap paradigm revealed that human infants in their first year of life show larger saccade latency under the overlap condition but not under the gap condition, as they were in an earlier stage of development (Farroni, Simion, Umiltà, & Barba, 1999; Hood & Atkinson, 1993; Matsuzawa & Shimojo, 1997). Autistic children also show the same slow saccade timing when compared with typically developing children when tested under the free viewing paradigm (Landry & Bryson, 2004) (the results are mixed when trained for this task (Goldberg et al., 2002; van der Geest, Kemner, Camfferman, Verbaten, & van Engeland, 2001)). Studies on macaque monkeys have shown that during free viewing of naturalistic dynamic scenes monkeys scan the scenes more rapidly and more widely than humans (Berg, et al., 2009; Shepherd, et al., 2010), which was similar to the apes in our studies. Thus, there is evidence that human infants and autistic children show the opposite pattern from apes and monkeys in their timing of gaze shift when compared with typically developed human adults.

Reduced eye viewing is commonly reported during face scanning in human infants (Hainline, 1978; Haith, Bergman, & Moore, 1977) and autistic individuals (Dalton et al., 2005; Klin, Jones, Schultz, Volkmar, & Cohen, 2002), although the results were somewhat controversial for autistic individuals (van der Geest, Kemner, Verbaten, & van Engeland, 2002). Monkeys view the eyes longer than other parts of the face (e.g., ears, mouth) (Dahl, Wallraven, Bulthoff, & Logothetis, 2009; Ghazanfar, Nielsen, & Logothetis, 2006; Gothard, Erickson, & Amaral, 2004; Guo, Robertson,

Mahmoodi, Tadmor, & Young, 2003; Keating & Keating, 1982; Mendelson, Haith, & Goldmanrakic, 1982; Nahm, et al., 1997), but somewhat more briefly than humans (Shepherd, et al., 2010). Thus, the evidence indicates that human infants and autistic children have a similar tendency to apes and macaques in their pattern of eye viewing when compared with typically developed human adults. Therefore, mosaic similarities and differences exist between nonhuman primates and humans of various developmental stages and clinical conditions. The direct causes of these mosaic similarities and differences are unknown. One suggestion from these patterns is that the eye movement characteristics of human and nonhuman primates do not necessarily represent the single strategies that are fully functional or immature/deficit, but rather represent alternative strategies that include both costs and benefits.

1.4.4. Remaining issues and future directions

There are several remaining issues concerning species comparisons and eye movement recordings. First, we used a passive viewing task for comparison between species. Although this paradigm is useful for examining the viewing patterns elicited by a subjects' spontaneous interest in the stimuli, it remains unclear what information the subjects retrieved from the stimuli. To specify such information more clearly, it is necessary to examine the viewing patterns while subjects are performing various tasks, in which the goal of the subjects is to solve a given task but not view the stimuli *per se*, e.g., eye movements while using multiple tools (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Land, 2006). Second, we have shown that there are several species differences between monkeys and humans and between apes and humans. However, it remains unclear how the eye movement characteristics of apes compare directly with those of macaque monkeys. Similarly, although we did not find particular differences among nonhuman great apes in their responses to faces, several studies have reported species differences in their social looking behavior, such as gaze following (Bräuer, Call, & Tomasello, 2005; Okamoto-Barth, Call, & Tomasello, 2007). Finally, we

examined the “reactive” aspects of eye movements in our studies. However, previous studies in humans have shown that the eyes move proactively while subjects are performing manual tasks (Johansson, Westling, Bäckström, & Flanagan, 2001; Land, 2006), and that the eyes move predictively to the other individual’s actions while subjects are viewing another individual performing manual tasks (Falck-Ytter, Gredebäck, & Von Hofsten, 2006; Flanagan & Johansson, 2003). Such proactive/predictive eye movements are largely unknown in nonhuman primates. Further studies are necessary to examine these issues.

1.4.5. Summary and conclusion

Using a recently developed video-based eye-tracker with a wide-angle lens, we recorded the eye movements of great apes noninvasively with head unrestraint even when apes and human experimenters were separated with transparent acrylic panels. Using a flexible experimental setup and calibration method, we recorded eye movements with a typical accuracy of less than 1°. We then directly compared apes and humans with regard to their free viewing patterns for scenes and faces under the same experimental conditions. We found striking similarities between species in overall eye movement patterns and responses to social stimuli, suggesting the homologous nature of eye movement functions in apes and humans. We also found quantitative differences between apes and humans that were not explained by differences in low-level or automatic eye movement functions. Thus, these results reflect species-specific behavioral strategies, which may reflect characteristic visual information processing and social communicative style. As comparative studies of ape eye movements have just begun, many issues remain unexplored in this paradigm. Among them, topics for future studies include goal-directed and proactive/predictive aspects of eye movement.

References

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge: MIT Press.

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *J Vis, 9*(5), 1-15.

Bräuer, J., Call, J., & Tomasello, M. (2005). All great ape species follow gaze to distant locations and around barriers. *J Comp Psychol, 119*(2), 145-154.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn Sci, 12*(5), 187-192.

Dahl, C. D., Wallraven, C., Bulthoff, H. H., & Logothetis, N. K. (2009). Humans and macaques employ similar face-processing strategies. *Curr Biol, 19*(6), 509-513.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nat Neurosci, 8*, 519-526.

De Waal, F. B. M. (1990). *Peacemaking among primates*. Cambridge: Harvard University Press.

Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. New York: Springer.

Ekman, P., & Friesen, W. V. (1978). *The facial action coding system (FACS): A technique for the measurement of facial action*. Palo Alto: Consulting Psychologists Press.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neurosci Biobehav Rev, 24*(6), 581-604.

Falck-Ytter, T., Gredebäck, G., & Von Hofsten, C. (2006). Infants predict other people's action goals. *Nat Neurosci, 9*(7), 878-879.

Farroni, T., Simion, F., Umiltà, C., & Barba, B. D. (1999). The gap effect in newborns. *Dev Sci, 2*(2), 174-186.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behav Brain Sci, 22*(04), 661-674.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behav Brain Sci, 16*(3), 553-567.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature, 424*(6950), 769-771.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*, 571-583.

Ghazanfar, A. A., Nielsen, K., & Logothetis, N. K. (2006). Eye movements of monkey observers viewing vocalizing conspecifics. *Cognition, 101*(3), 515-529.

Goldberg, M. C., Lasker, A. G., Zee, D. S., Garth, E., Tien, A., & Landa, R. J. (2002). Deficits in the initiation of eye movements in the absence of a visual target in adolescents with high functioning autism. *Neuropsychologia, 40*(12), 2039-2049.

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge: Belknap Press.

Gothard, K. M., Erickson, C. A., & Amaral, D. G. (2004). How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Anim Cogn, 7*(1), 25-36.

Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y., & Young, M. P. (2003). How do monkeys view faces?—a study of eye movements. *Exp Brain Res, 150*(3), 363-374.

Hainline, L. (1978). Developmental changes in visual scanning of face and nonface patterns by infants. *J Exp Child Psychol, 25*(1), 90-115.

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. Science, *198*(4319), 853-855.

Hattori, Y., Kano, F., & Tomonaga, M. (2010). Differential sensitivity to conspecific and allospecific cues in chimpanzees and humans: A comparative eye-tracking study. *Biology Lett, 6*(5), 610-613.

Hayhoe, M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *J Vis, 3*(1), 49-63.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends Cogn Sci, 7*(11), 498-504.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annu Rev Psychol, 50*, 243-271.

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *J Exp Psychol Hum Percept Perform, 25*(1), 210-228.

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Res, 45*(13), 1707-1724.

Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., & Fujita, S. (2010). Facial perception of conspecifics: Chimpanzees (*Pan troglodytes*) preferentially attend to proper orientation and open eyes. *Anim Cogn, 13*(5), 679-688.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behav Dev, 16*(4), 405-422.

Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychol Sci, 19*(10), 998-1006.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res, 40*(10-12), 1489-1506.

Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001). Eye-hand coordination in

object manipulation. *J Neurosci, 21*(17), 6917-6932.

Kano, F., Hirata, S., Call, J., & Tomonaga, M. (2011). The visual strategy specific to humans among hominids: A study using the gap-overlap paradigm. Vision Res., 51, 2348-2355..

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: a comparative eye-tracking study. *Proc Roy Soc B, 276*(1664), 1949-1955.

Kano, F., & Tomonaga, M. (2010). Face scanning in chimpanzees and humans: Continuity and discontinuity. *Anim Behav, 79*(1), 227-235.

Kano, F., & Tomonaga, M. (2011a). Perceptual mechanism underlying gaze guidance in chimpanzees and humans. *Anim Cogn, 14*(3), 377-386.

Kano, F., & Tomonaga, M. (2011b). Species difference in the timing of gaze movement between chimpanzees and humans. *Anim Cogn, 14*(6), 879-892.

Keating, C. F., & Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception, 11*(2), 211-219.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch Gen Psychiatry, 59*(9), 809-816.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J Hum Evol, 40*(5), 419-435.

Land, M. F. (2006). Eye movements and the control of actions in everyday life. *Prog Retin Eye Res, 25*(3), 296-324.

Landry, R., & Bryson, S. E. (2004). Impaired disengagement of attention in young children with autism. *J Child Psychol Psychiat, 45*(6), 1115-1122.

Malcolm, G. L., Lanyon, L. J., Fugard, A. J. B., & Barton, J. J. S. (2008). Scan patterns during the

processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects. *J Vis, 8*(8), 1-9.

Matsuzawa, M., & Shimojo, S. (1997). Infants'fast saccades in the gap paradigm and development of visual attention. *Infant Behav Dev, 20*(4), 449-455.

Mendelson, M. J., Haith, M. M., & Goldmanrakic, P. S. (1982). Face scanning and responsiveness to social cues in infant rhesus monkeys. *Dev Psychol, 18*(2), 222-228.

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nat Rev Neurosci, 5*(3), 218-228.

Nahm, F. K. D., Perret, A., Amaral, D. G., & Albright, T. D. (1997). How do monkeys look at faces? *J Cogn Neurosci, 9*(5), 611-623.

Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychol Sci, 18*(5), 462-468.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychol Bullet, 124*, 372-422.

Robinson, D. A. (1968). Eye movement control in primates. *Science, 161*(3847), 1219-1224.

Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr Biol, 20*(7), 649-656.

Stacey, P. C., Walker, S., & Underwood, J. D. M. (2005). Face processing and familiarity: Evidence from eye-movement data. *Br J Psychol, 96*(4), 407-422.

Thomsen, C. E. (1974). Eye contact by non-human primates toward a human observer. *Anim Behav, 22*, 144-149.

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *J Hum Evol,*

52(3), 314-320.

van der Geest, J. N., Kemner, C., Camfferman, G., Verbaten, M. N., & van Engeland, H. (2001). Eye movements, visual attention, and autism: A saccadic reaction time study using the gap and overlap paradigm. *Biol Psychiatry, 50*(8), 614-619.

van der Geest, J. N., Kemner, C., Verbaten, M. N., & van Engeland, H. (2002). Gaze behavior of children with pervasive developmental disorder toward human faces: A fixation time study. *J Child Psychol Psychiat, 43*(5), 669-678.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6*(3), 313.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395-1407.

Yarbus, A. L. (1967). Eye movements during perception of complex objects *Eye Movements and Vision* (pp. 171-211). New York: Plenum Press.

Figure Captions

Figure 1-1. A chimpanzee on an eye-tracker.

Figure 1-2. A schematic representation of retinal (pupil) and corneal reflections. The cross points of the solid and dotted lines indicate the centroid of retinal and corneal reflections, respectively.

Figure 1-3. Examples of eye movement paths for a chimpanzee and a human subject during 3-s free viewing of the presented scene (toned down for clarity). The circles represent fixations (scaled in size to their durations), and the lines represent saccades. Fixation was scored if the gaze remained stationary within a radius of 50 pixels (approx. 1.3° at a viewing distance of 60 cm) for 75 ms (more than five 60-Hz recording samples).

Figure 1-4. Examples of spatial distribution of fixation by six chimpanzees and humans during 3-s free viewing of the presented scene (toned down for clarity).

Figure 1-5. Examples of spatial distribution of fixations by six chimpanzees and humans during 3-s free viewing of the presented scene (toned down for clarity). Also shown is the low-level saliency map generated from and superimposed on the presented scene. Dark areas indicate areas of high saliency.

Figure 1-6. (**a**) Average fixation duration for chimpanzees (n = 6) and humans (n = 18) during 3-s free viewing of the four different scenes. Six exemplars were prepared for each type of scene (24 exemplars in total). (**b)** Frequency distribution for the fixation duration (ms); data in (**a)** were pooled

for each species.

Figure 1-7. (**a)** Gap-overlap paradigm (free viewing). Twenty-four trials were conducted in total for each condition. (**b)** Frequency distribution of saccade latency (ms) under the gap and overlap conditions.

Figure 1-8. Sequence from a movie presented in our study (free viewing). Forty-eight trials were conducted in total. Also shown is the number of fixated images (the number of images on which subjects fixated at least once) and total viewing time (ms) for the images (the sum of the fixation durations on any image).

Figure 1-9. Examples of eye movement paths of six chimpanzees and humans during 2-s free viewing of the presented faces (toned down for clarity).

Figure 1-10. Examples of eye movement paths of a chimpanzee and a human subject during 2-s free viewing of the presented facial expressions (toned down for clarity).

Figure 1-11. Examples of eye movement paths of a chimpanzee and a human subject during 3-s free viewing of the presented chimpanzee or human model (toned down for clarity).

Figures

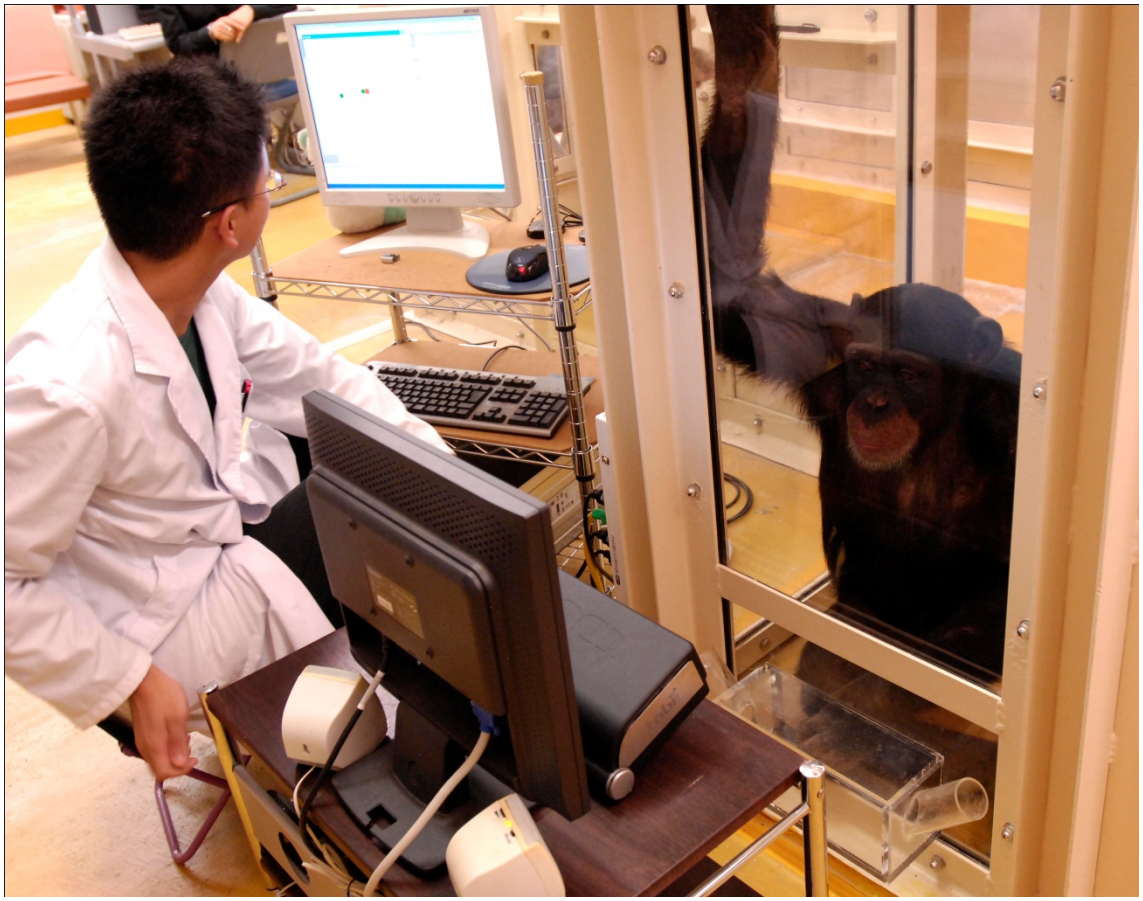


Figure 1-1

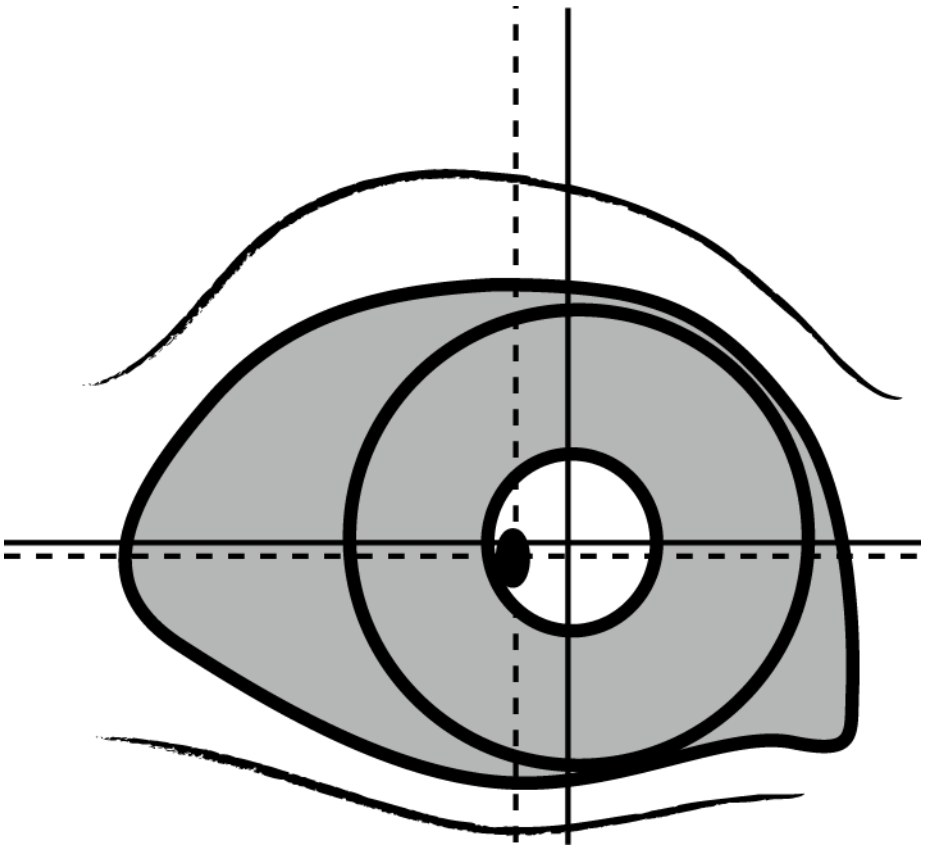


Figure 1-2

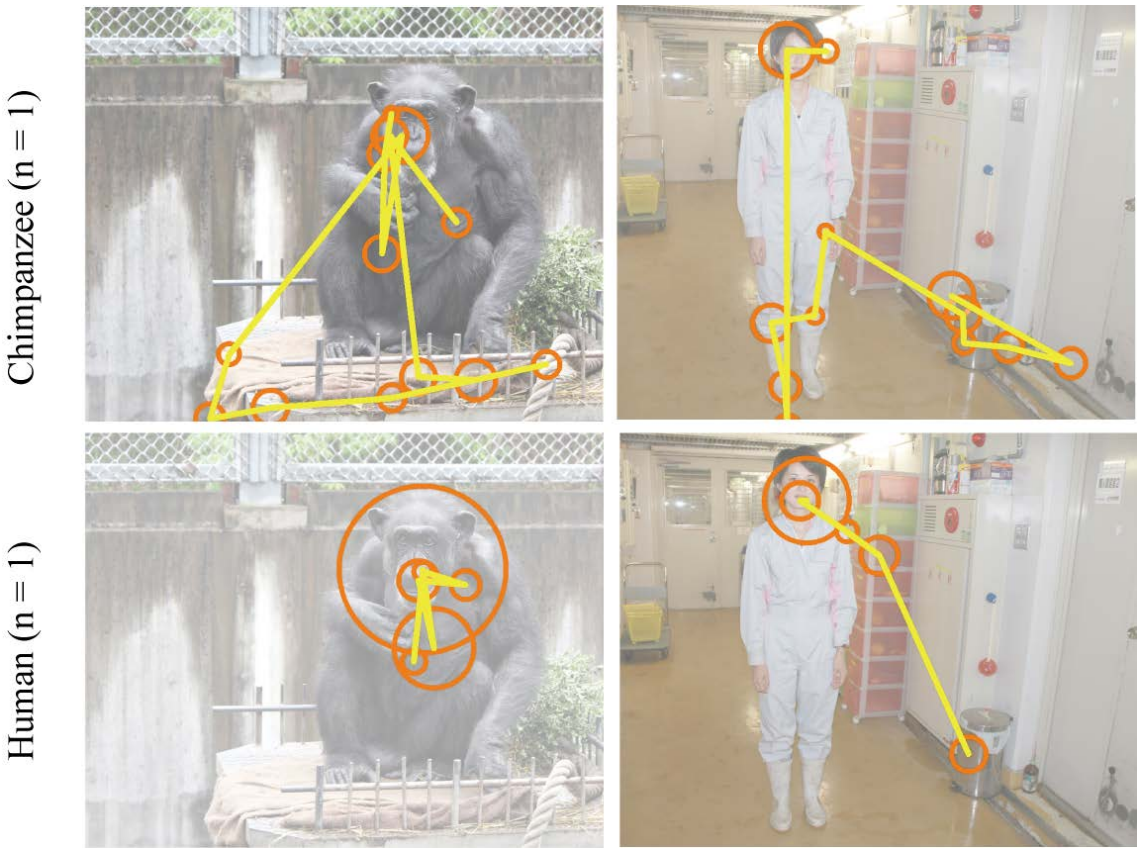


Figure 1-3

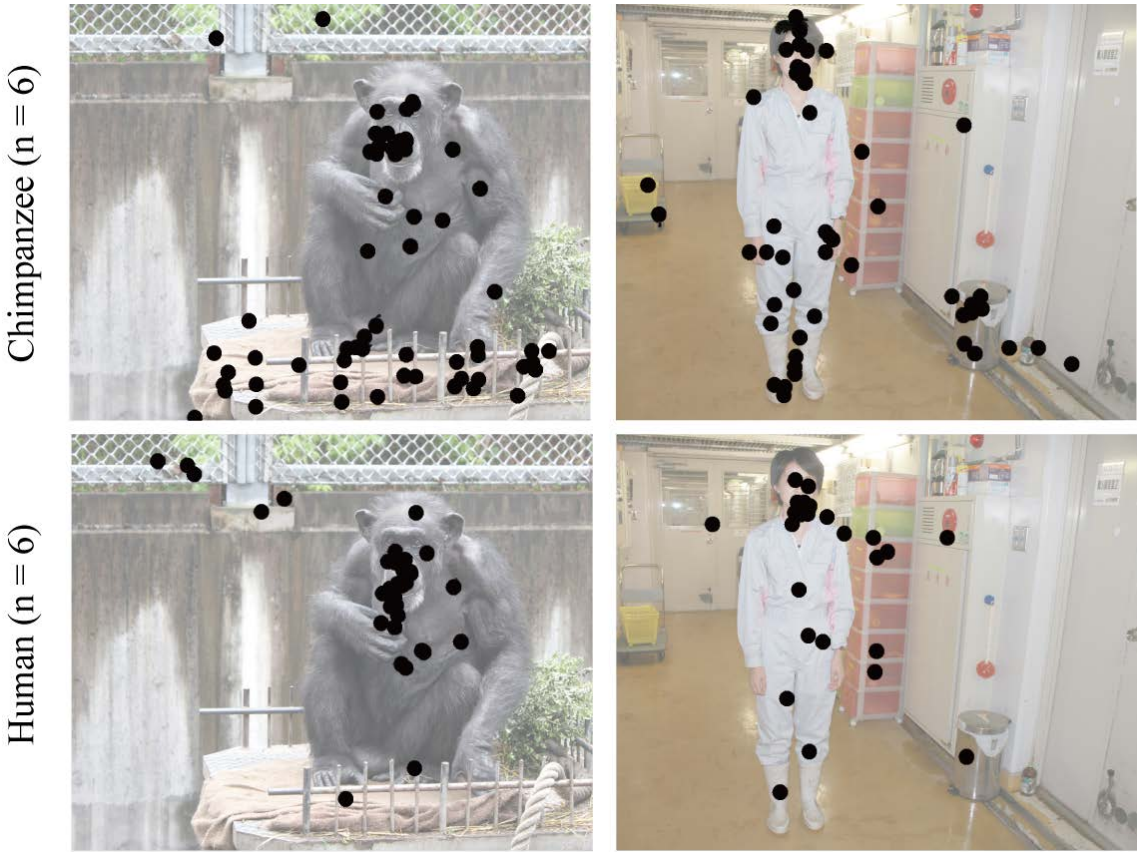


Figure 1-4

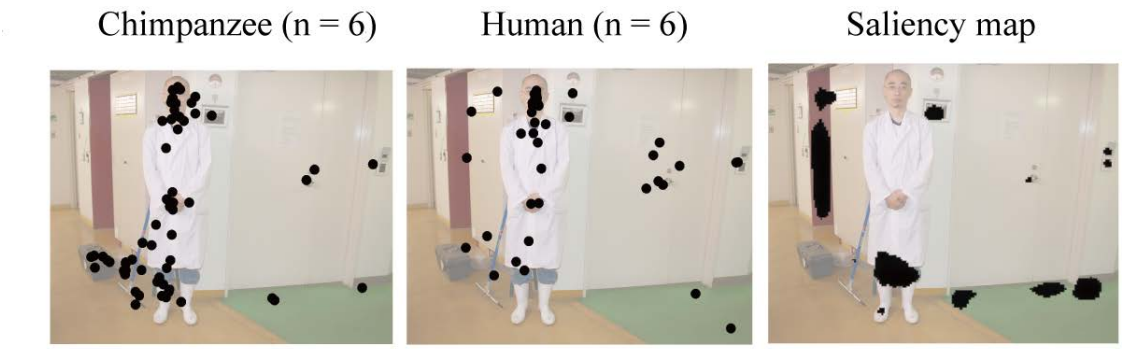
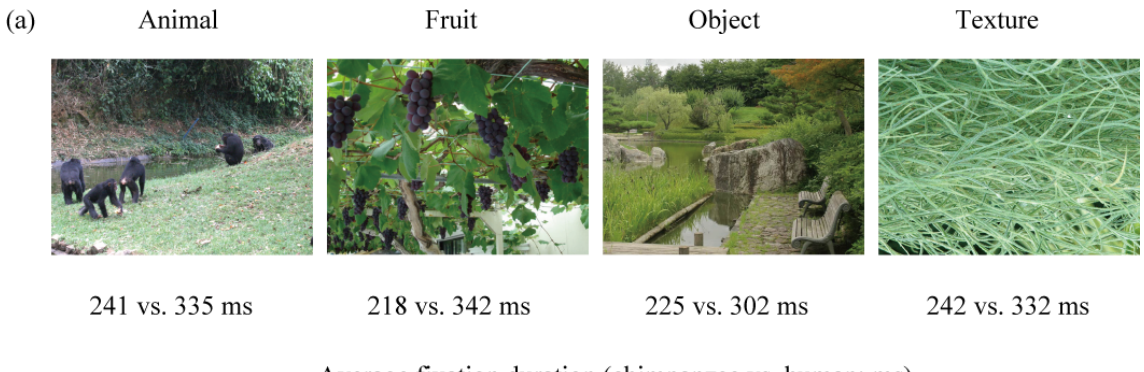


Figure 1-5


(a)
Animal
Fruit
Object
Texture
241 vs. 335 ms
218 vs. 342 ms
225 vs. 302 ms
242 vs. 332 ms
Average fixation duration (chimpanzee vs. human; ms)

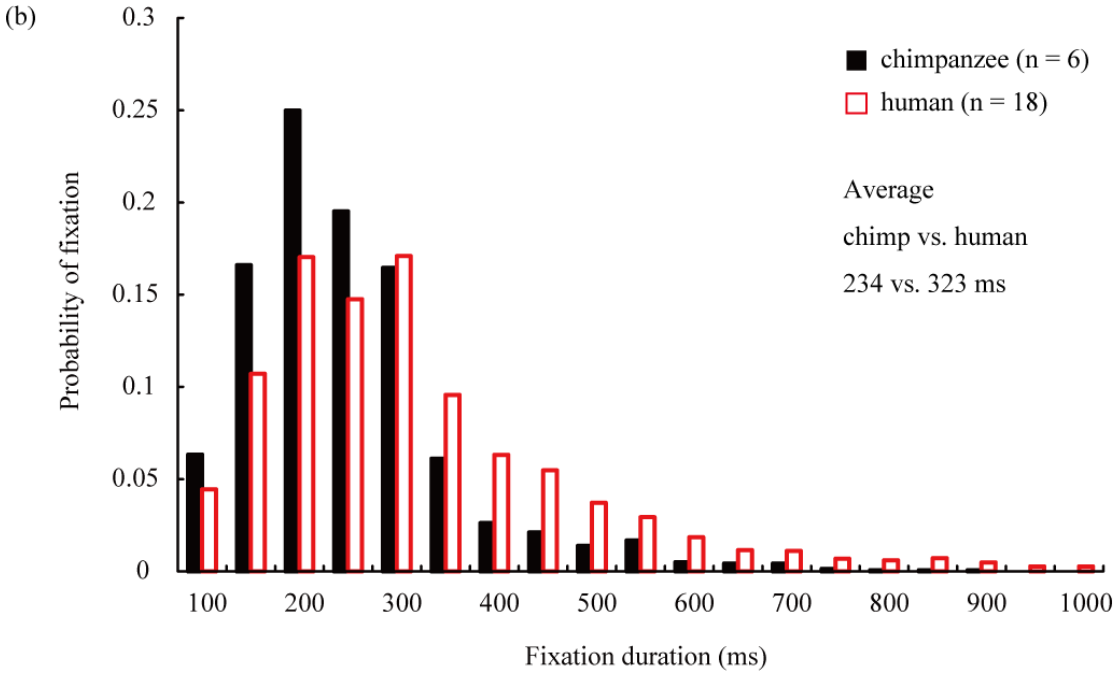

(b)
Probability of fixation
0.3
0.25
0.2
0.15
0.1
0.05
0
100
200
300
400
500
600
700
800
900
1000
Fixation duration (ms)
chimpanzee (n = 6)
human (n = 18)
Average
chimp vs. human
234 vs. 323 ms

Figure 1-6

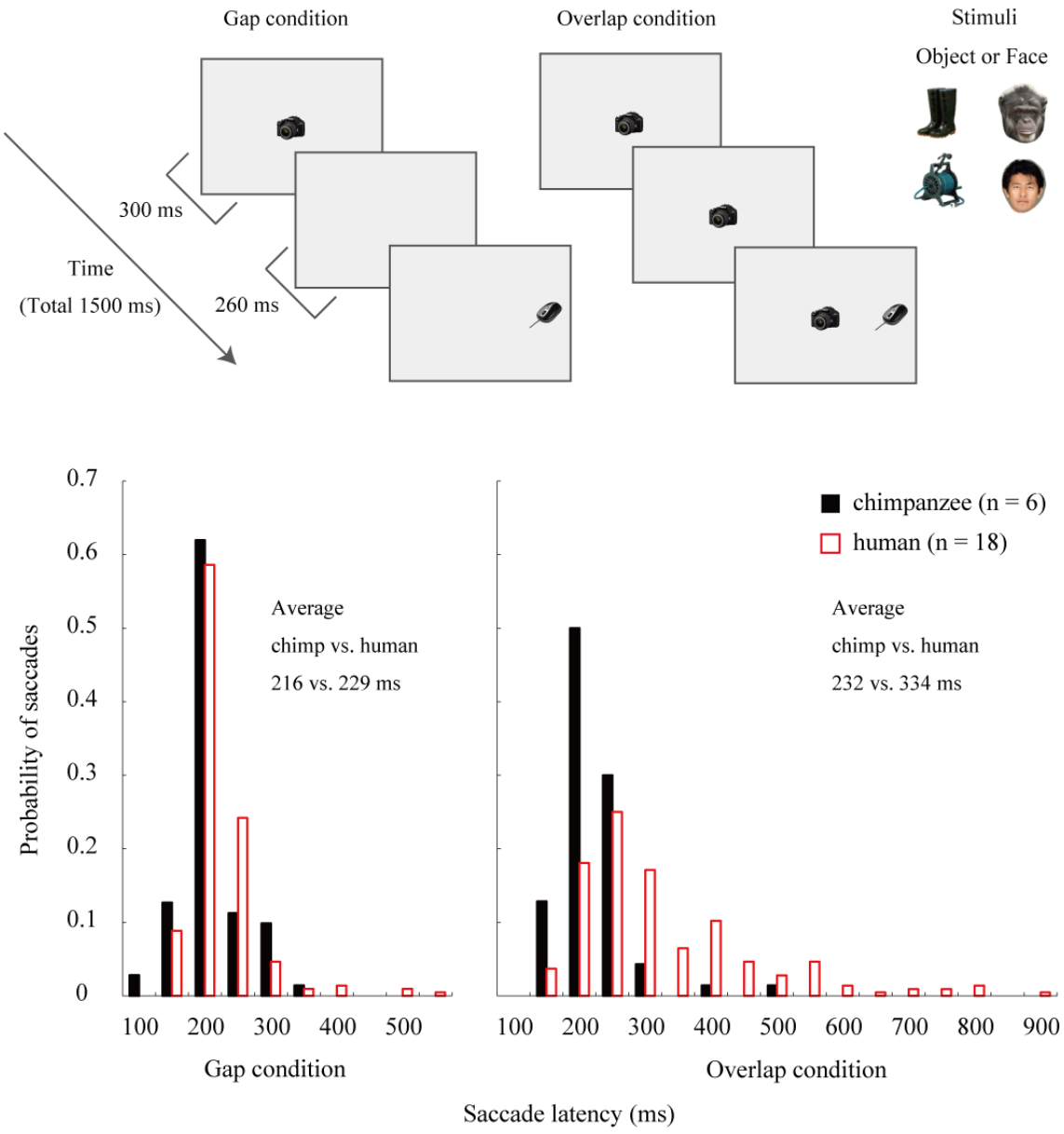
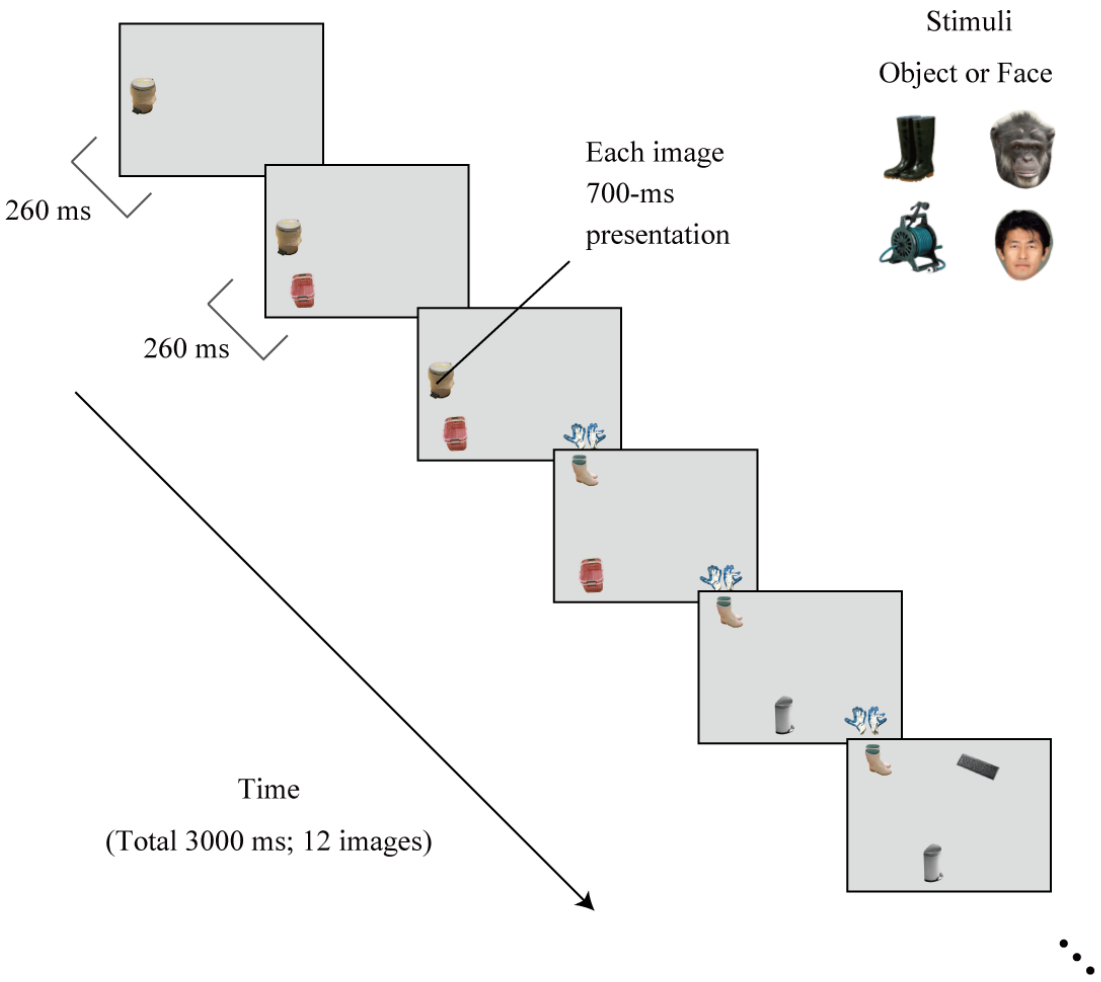

Gap condition
Overlap condition
Stimuli
Object or Face
300 ms
260 ms
Time
(Total 1500 ms)
chimpanzee (n = 6)
human (n = 18)
Average
chimp vs. human
216 vs. 229 ms
Average
chimp vs. human
232 vs. 334 ms
Probability of saccades
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
100 200 300 400 500
Gap condition
100 200 300 400 500 600 700 800 900
Overlap condition
Saccade latency (ms)

Figure 1-7



	Number of fixated images	Total viewing time (ms)
Chimpanzee (n = 6)	9.3	2,051
Human (n = 18)	6.1	1,498

Figure 1-8

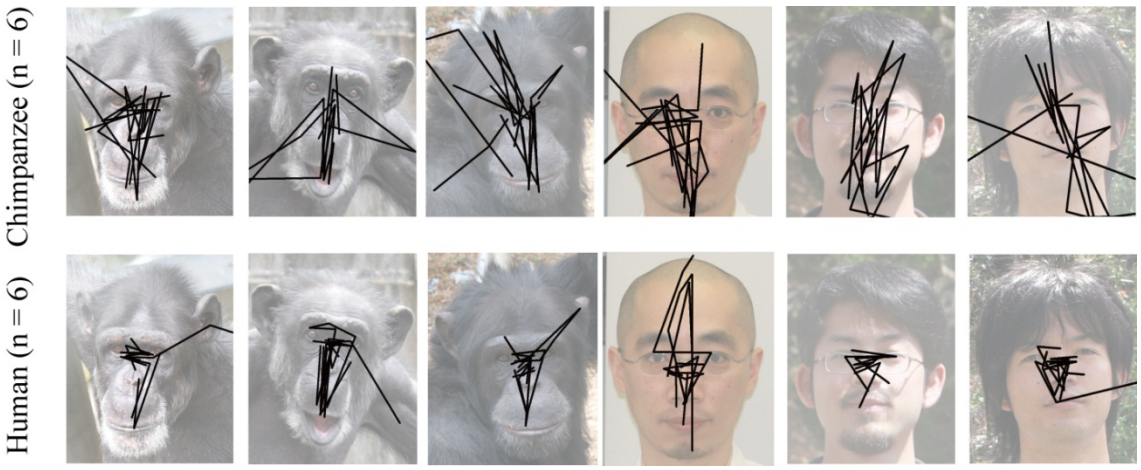


Figure 1-9

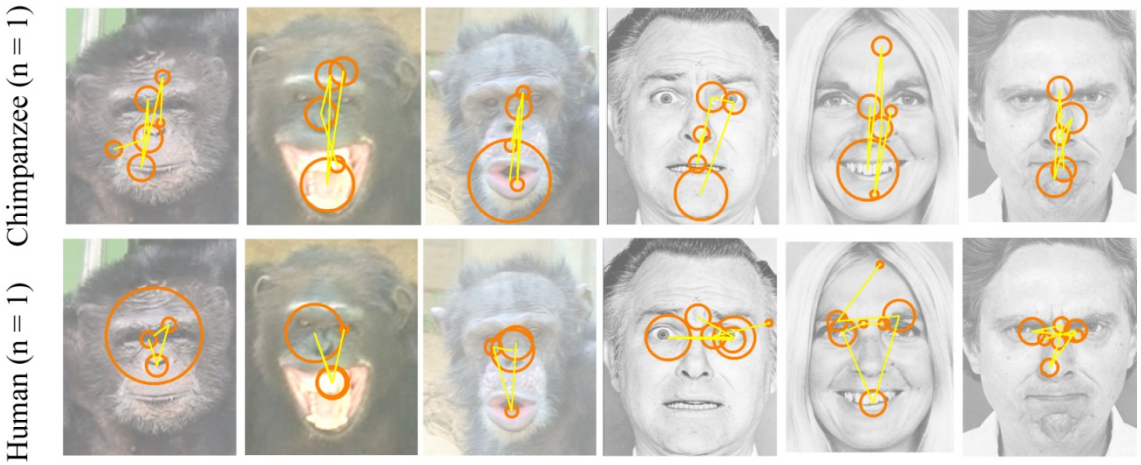


Figure 1-10

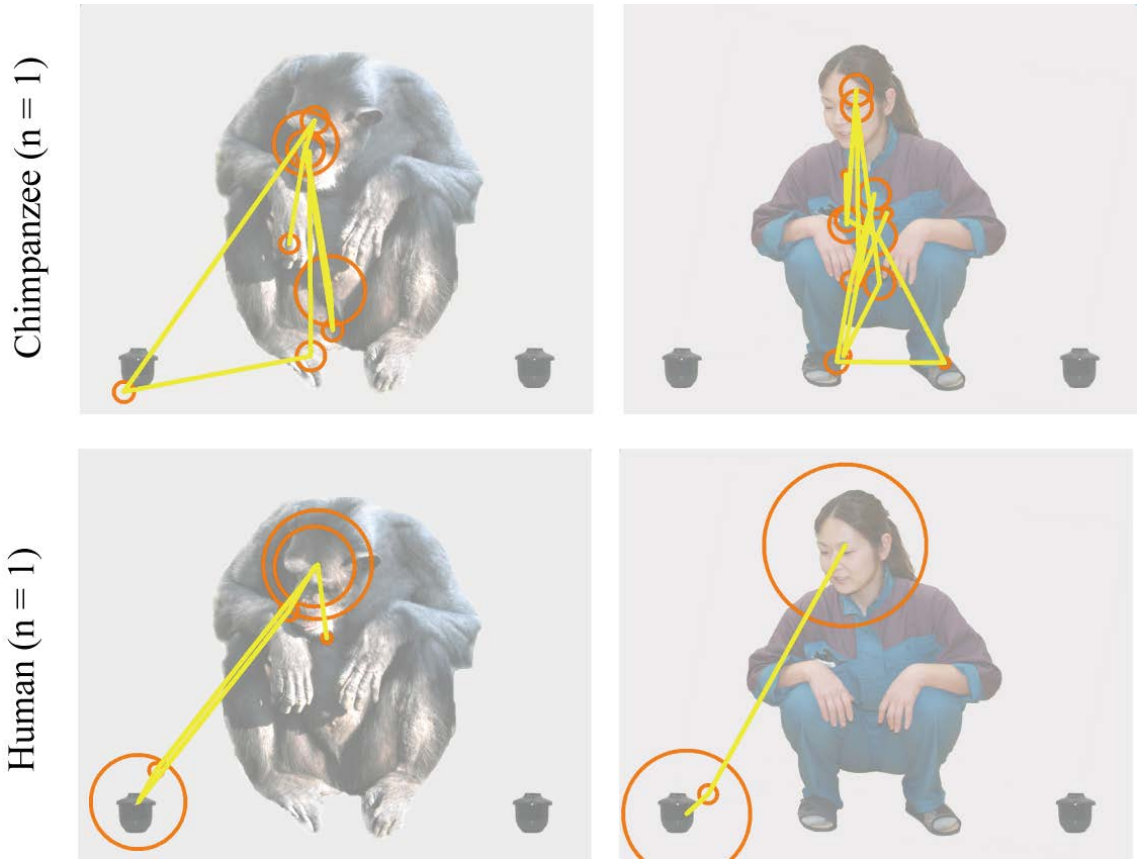


Figure 1-11

Chapter 2: Specific Researches

Chapter 2-1

How chimpanzees look at pictures: a comparative eye-tracking study

Introduction

Eye tracking enables the direct assessment of eye movements and has been an important method in studies of both humans and non-human primates, particularly monkey species. Eye movements can potentially reveal a variety of cognitive and emotional processes, from visual-spatial attention to social information processing and motivational change. Several comparative cognitive eye-tracking studies have revealed similarities in the eye movements of monkeys and humans (Gothard et al. 2004; Keating & Keating 1982; Nahm et al. 1997). However, these studies only indirectly compared the eye movements of these species. Very little research has made direct comparisons under similar conditions (but see (Kobayashi & Kohshima 2001)). Thus, little is currently known about the similarities and differences in eye-movement patterns of non-human and human primates. Furthermore, despite the critical importance of ape species in comparative-cognitive science (Matsuzawa et al. 2006), no studies have compared eye-tracking between apes and humans. In light of these issues, we studied eye-tracking in humans and our most closely related species, the chimpanzee (*Pan troglodytes*).

Early studies of eye-tracking have noted that human gaze control is highly regular; viewers tend to concentrate their fixations on semantically informative regions when shown pictures of scenes or faces (Buswell 1935; Yarbus 1967). For example, when viewing a picture containing human figures, humans were more likely to fixate on human figures compared to the background, on faces compared to other parts of the body, and on the eyes compared to other facial features. Subsequently, studies have repeatedly demonstrated that the human gaze is drawn not only to visually salient regions in terms of color, contrast, and edge orientation, but also to interesting or meaningful regions based on the viewer's knowledge (Henderson & Hollingworth 1999a). This active, top-down guidance of picture viewing has also been observed in monkeys and chimpanzees. In eye-tracking studies, monkeys intensively fixated on the eye region when viewing images of the

faces of conspecifics (Gothard et al. 2004; Keating & Keating 1982; Nahm et al. 1997), as has been observed in humans. In other studies, chimpanzees allowed to choose various pictures of scenes by pressing buttons or levers preferentially selected pictures containing human figures (Fujita & Matsuzawa 1986; Tanaka 2003). In a study using the same procedure, monkeys exhibited visual preferences for pictures containing the face region compared to those featuring various other body parts, suggesting the importance of the face in body viewing ((Fujita 1993) also see (Tomonaga 1994)). These observations suggest similarities in viewing behavior among monkeys, apes, and humans when looking at pictures containing scenes, bodies, and faces. However, information on both similarities and differences in viewing patterns of primates is important to more specifically delineate evolutionary elements of gaze behavior. Therefore, we made a direct, quantitative comparison of eye movements between two primate species using eye-tracking methodology.

The purpose of our study was to expand on the above-mentioned research by applying a comparative eye-tracking method to chimpanzees and humans. Of particular interest was the direct comparison of scene-viewing by the two species. In our experiment, both chimpanzee and human participants were tested using the same experimental conditions and were presented with identical pictures portraying chimpanzees, humans, and non-primate mammals. To promote natural gaze behavior, the pictures consisted of complex scenes, that is, full-body images of animals against naturalistic backgrounds. We predicted that, in accordance with previous human studies, both chimpanzees and humans would look at the informative regions disproportionately longer compared to other features of the pictures; specifically, they would look at the animal figures longer than at the background and at the face regions longer than at other parts of the body. A second goal of our study was to observe how the two species scan images of the body. The face was of particular interest, because this region is highly informative for these social primates. Studies of non-verbal communication suggest that the face conveys a wealth of information, such as identity, sex, and

emotion in both species (Argyle 1967; Van Hooff 1967).

In addition to the traditional index of looking behavior, looking duration, we also examined the sequence of fixations as participants viewed the pictures. In human eye tracking, the face image within a scene is detected immediately after picture onset, suggesting rapid processing of person information (Fletcher-Watson et al. 2008). The same phenomenon was confirmed in a visual search task (Hershler & Hochstein 2005). We replicated this visual search task in a related study and observed the rapid detection of the face of conspecifics in chimpanzees (Tomonaga 2009). In this related study, the three adult chimpanzees showed quicker response time (RT) to faces than to artificial object targets. The present study addressed the same issue using an eye-tracking paradigm under free-viewing conditions. We predicted that first fixations were more likely to be located on the face than on other features of the pictures. We also examined the effect of prior visual experiences on face detection. The chimpanzee and human participants had less (or no) experience observing non-primate mammals than viewing their own or closely related species. The goal of this comparison was to determine how patterns of fixation on faces when viewing animals with familiar morphology are generalized when viewing animals with unfamiliar morphology.

Methods

Participants.

Six chimpanzees (*Pan troglodytes*; one male juvenile, two female juveniles, and three female adults) and 21 humans (five males, 16 females; all adults) participated in this study. The chimpanzees live within a social group of 14 individuals in an environmentally enriched outdoor compound and an attached indoor residence (Matsuzawa 2006). The chimpanzees have been familiar with humans since birth. The care and use of the chimpanzees adhered to the 2002 version of the Guidelines for the Care and Use of Laboratory Primates by the Primate Research Institute, Kyoto University. The experimental protocol was approved by the Animal Welfare and Care Committee of the same institute. The human participants were graduate and undergraduate students (all Japanese), who participated in the experiment voluntarily. Thirteen out of the 21 humans had extensive experience with observations of chimpanzees.

Apparatus.

The chimpanzee and human participants used exactly the same apparatus to allow for direct comparisons. A corneal reflection technique was used to measure eye movements. We used a table-mounted eye tracker (60 Hz; Tobii x120, Tobii Technology AB) with a wide-angle lens (± 40 degree in the semi-circle above camera) which allowed for relatively large head movements of participants during tracking and also obviated the necessity to restrain them. All participants sat in an experimental booth (2.5 m wide × 2.5 m deep × 2.1 m high), with the experimenter and the participants separated by transparent acrylic panels. The participants viewed a 17-inch LCD display (1280 × 1024 pixels), and the eye tracker captured their eye movements through the acrylic panels. The eye tracker and the display were mounted on a movable platform (0.6 × 0.6 × 0.4 m). The distance between the platform and the participants was adjusted to the point at which the gaze was most accurately recorded (approximately 60 cm). One degree of gaze angle corresponded to

approximately 1 cm on the screen at a 60-cm viewing distance.

Although the eye morphologies of the two species are somewhat different (Kobayashi & Kohshima 2001), the accuracy of gaze data was quite similar. In preliminary recordings, we confirmed that the average error when viewing the screen (the distance between measured and intended gaze points) was < 0.5° in both species. Although the eye tracker sometimes lost the participants' eyes because of postural changes or eye blinks, these values were similar for both species (chimpanzees: 5.9 ± 1.3%; humans: 7.1 ± 1.5% in 3-s presentations (± SEM)). Thus, no special corrections of the raw tracking data were necessary in this study.

Stimuli.

Non-primate mammals (hereafter, "mammal"), chimpanzees, and humans were used as picture models. The pictures consisted of single, foregrounded, full-body images of an individual with a naturalistic, complex background (See Figure 2-1-1 for examples of the pictures shown). We prepared 24 pictures of mammals representing 19 species that are typically seen at zoos (e.g., bears, rhinoceros, giraffes, and elephants). We also prepared 24 chimpanzee pictures and 24 human pictures of 12 individuals each (six each of individuals familiar and unfamiliar to all chimpanzee participants and to 11 out of 21 human participants). The pictures were manipulated so that the variance of the size of animal models was small. Pictures were then converted to 1000 × 800 pixels with surrounding gray-colored frames (1280 × 1024 pixels in total).

Procedure.

Before presenting the pictures, we conducted habituation training of the participants and calibration of the eye tracker to obtain accurate gaze recordings. Both species followed nearly the same procedure with some exceptions. Below, the procedure for chimpanzees is described first.

Training and calibration. In the initial training, chimpanzees were required to face the screen so that their corneal reflections in both eyes were captured by the eye tracker. The trial succeeded if they

maintained this posture for 1.5 s. After being well-habituated to this procedure, a two-point calibration was conducted for each chimpanzee. To attract their attention, a small video clip (200 × 200 pixels) was presented at each calibration point. Once successfully calibrated, the chimpanzees were required to look at the fixation point (a red square), which appeared at a random position on the screen. The purpose of this training was to evaluate calibration accuracy. The trial succeeded if the participant fixated on the fixation spot for at least 1 s within a distance of 3°. The calibration was repeated if it was not initially accurate enough. The chimpanzees were rewarded with a piece of apple for successful trials. The performance criterion to advance to the next training/calibration phase was set to 90% of trials. All chimpanzees finished the entire procedure within two weeks of daily sessions of 10–15 minutes each.

Testing.

The trial was initiated by the subject's looking at a fixation point that appeared at a random position on the screen. If the chimpanzees held the fixation position for 250 ms, a picture was presented. Once the picture appeared, they could freely move their eyes to look at the picture. Each picture was presented for 3 s. The chimpanzees were rewarded after the presentation regardless of their viewing behaviors. The purpose of this reward was to maintain the chimpanzees' motivation to participate in the testing. The entire testing of chimpanzees was conducted over a 10 day period. A daily session lasted for 5–10 minutes with presentations of only seven or eight pictures to keep their interests. These pictures were randomly drawn from each stimulus group. Each chimpanzee viewed 72 pictures in total. Before testing on each day, we evaluated the calibration record of the individual on five fixation spots and conducted a recalibration if necessary.

The procedural differences between species were as follows: (1) no reward was given to humans during the experiment, (2) humans were verbally instructed to view the pictures as they normally would, and (3) all training and testing in humans were conducted within a single day and

lasted for 20–30 minutes.

Data analysis.

We divided each picture into several features (areas of interest; AOI) to quantitatively analyze participants' viewing patterns. Two scene features were defined: (1) background and (2) entire body. For more detailed analysis, entire body was divided into (3) face and (4) other parts of the body. The other parts of the body region was further divided into three regions: (5) torso, (6) arms, and (7) legs. For mammal pictures, the arms and legs regions were combined into (8) limbs. Each feature was mapped by outlining it with a polygon (see Figure 2-1-2 for examples). To avoid errors in gaze estimation, the polygon was drawn slightly larger than the actual outline (approximately 20 pixels on the edge). The recorded eye-tracking samples were added to the AOI region if they were within the polygon. If two or more AOIs were duplicated, the samples were added to the upper AOI.

We used five dependent variables indicating attention: gaze time (the sum of fixation duration), proportion of fixation (the proportion of pictures in which any AOI was the target of the fixation), fixation number, fixation duration, and distance between fixations (saccade (rapid eye movement that shifts the gaze from one fixation to another) size). A fixation was scored if the gaze remained stationary (within a radius of 50 pixels) for at least 75 ms (more than five measurement samples). Otherwise, the recorded sample was defined as part of a saccade. In order to limit analysis to the visual information actually available to the participants, we excluded the samples recorded during the first 200 ms, thereby eliminating fixations that followed the offset of the fixation spot. We also excluded samples recorded during saccades.

For statistical analyses, we distinguished between within-species and between-species comparisons. For within-species comparisons, we tested for differences in viewing patterns over scene/body features within each species. For between-species comparisons, we tested for

interactions between species and scene/body features in viewing patterns. These comparisons were tested using repeated-measures analysis of variance (ANOVA) or paired t-tests in SPSS 13.0. Greenhouse-Geisser's epsilon was used for conservative adjustments to the degrees of freedom when sphericity did not hold. Post-hoc comparisons were conducted using Dunnett tests for within-species comparisons and t-tests with Bonferroni correction for between-species comparisons. Alpha was set at $p < 0.05$ for all analyses, although Bonferroni correction adjusted the alpha level for the number of between-species comparisons of scene/body features.

Results

Typical scanpaths of chimpanzees and humans illustrate that the two species exhibited similar but distinctive patterns of eye movement (Figure 2-1-1). Eye-movement patterns were quantified by dividing each picture into several scene/body features, as follows.

Figure 2-1-1

Gaze time.

Consistent with our prediction, both chimpanzee and human participants looked at the entire body longer than at the background region when viewing pictures of mammals (t_5 = 6.19, p = 0.002; t_{20} = 23.1, p < 0.001, respectively), chimpanzees (t_5 = 4.51, p = 0.006; t_{20} = 20.5, p < 0.001, respectively) and humans (t_5 = 1.82, p = 0.128 (n.s.; see next paragraph for this exception) ; t_{20} = 9.30, p < 0.001, respectively; Figure 2-1-2). When we focus on the entire body, we confirmed a non-random pattern in gaze time on each body feature (face, torso, etc.) by both chimpanzee and human participants when viewing pictures of mammals ($F_{2, 10}$ = 51.2, p = 0.003; $F_{1.2, 24}$ = 733, p < 0.001, respectively), chimpanzees ($F_{2,10}$ = 33.6, p < 0.001; F1.4, 28 = 584, p < 0.001, respectively), and humans ($F_{2, 10}$ = 13.5, p = 0.014; $F_{1.1, 22}$ = 153, p < 0.001, respectively; Figure 2-1-2). Post-hoc tests (Dunnett tests with the face region as a control) revealed that the face was the most extensively inspected region among all body features in both species, which was consistent with our predictions.

Figure 2-1-2

The sum of gaze times may not be equal in chimpanzees and humans for two reasons. First, chimpanzees made more saccades than humans (see below), and the gaze time did not include these

saccades (see Methods). Second, chimpanzees sometimes glanced outside of the pictures (120 ms on average), while humans rarely did so. As an interesting exception, one chimpanzee participant looked at the entire body for a shorter period than at the background region when viewing pictures of humans (707 vs. 1649 ms), which was opposite to the tendency of the other participants. The other five chimpanzees viewed the entire body longer than the background (1585 ± 123 ms vs. 562 ± 120 ms (mean ± SEM); $t_4 = 4.26$; $p = 0.013$). To examine the effect of model familiarity on gaze time in chimpanzee/human pictures, we tested the interactions between model familiarity and scene features (background, face and other parts of body) for each species. However, no significant interaction was found either in chimpanzee or human participants (N = 6, N = 11, respectively; see "Stimuli" for details) when viewing pictures of chimpanzees ($F_{2, 10} = 1.86$, $p = 0.20$; $F_{2, 20} = 1.93$, $p = 0.17$, respectively) and humans ($F_{2, 10} = 0.79$, $p = 0.48$; $F_{2, 20} = 0.28$, $p = 0.75$, respectively). Thus, this factor was not further considered in this study.

In the between-species comparisons, a significant interaction was found between species and scene features (entire body, background) when participants were shown pictures of mammals ($F_{1.1, 29} = 19.8$, $p < 0.001$), chimpanzees ($F_{1.0, 27} = 9.29$, $p = 0.004$), and humans ($F_{1.0, 25} = 5.22$, $p = 0.03$; Figure 2-1-2). Similarly, the interactions between species and body feature (face, torso, etc.) were also significant when subjects were shown pictures of mammals ($F_{1.2, 31} = 13.1$, $p < 0.001$), chimpanzees ($F_{1.1, 29} = 31.8$, $p < 0.001$), and humans ($F_{1.5, 39} = 10.8$, $p < 0.001$; Figure 2-1-2). Post-hoc tests (t-tests with Bonferroni correction) revealed that human participants looked at the face region longer compared to chimpanzees when viewing any pictures. On the other hand, chimpanzees looked at some parts of the body longer compared to humans, for example, the torso and legs in pictures of chimpanzees and the arms and legs in pictures of humans. Although humans looked at the entire body region longer compared to chimpanzees, this difference can be attributed to their longer viewing of the face region (Figure 2-1-2).

Number of fixations, fixation duration, and distance between fixations.

Since we were primarily interested in general differences in these variables between species, these analyses only involve between-species comparison. To avoid redundancy, all pictures were pooled, and torso, arms, and legs (or limbs) regions were combined into other parts of the body. In general, chimpanzees exhibited a greater number of fixations ($F_{1, 25}$ = 278, p < 0.001), longer fixation durations ($F_{1, 25}$ = 289, p < 0.001), and longer distances between fixations (t_{25} = 6.54, p < 0.001) compared to humans (Figure 2-1-3). These results indicate that chimpanzees shifted their fixation location more quickly (i.e., showed more saccades) and more broadly than did humans. There were significant interactions between species and scene/body features in number of fixations ($F_{1.3, 34}$ = 7.34, p = 0.006) and fixation duration ($F_{1.1, 28}$ = 17.0, p < 0.001; Figure 2-1-3). Post-hoc tests (t-tests with Bonferroni correction) confirmed that, compared to humans, chimpanzees exhibited a greater number of fixations on other parts of the body (t_{25} = 5.84, p < 0.001) and on the background region (t_{25} = 3.15, p = 0.004), but not on the face (t_{25} = 1.32, p = 0.198; Figure 2-1-4). In addition, chimpanzees exhibited shorter fixation durations on the face (t_{25} = 4.52, p < 0.001) and on other parts of the body (t_{25} = 2.57, p = 0.016), but not on the background (t_{25} = 1.92, p = 0.065; Figure 2-1-3) compared to humans.

Figure 2-1-3

Proportion of fixations.

We confirmed a non-random pattern in the first fixation on each scene/body feature by both chimpanzee and human participants when shown pictures of mammals ($F_{3, 15}$ = 19.1, p < 0.001; $F_{1.1, 29}$ = 92.6, p < 0.001, respectively), chimpanzees ($F_{1.5, 7.7}$ = 14.0, p = 0.004; $F_{1.2, 25}$ = 292, p < 0.001, respectively), or humans ($F_{1.2, 6.1}$ = 11.7, p = 0.012; $F_{1.0, 21}$ = 89.2, p < 0.001, respectively;

Figure 2-1-4a). Post-hoc comparisons (Dunnett tests with face region as a control) revealed that both species were most likely to attend to the face region during the first fixation when shown pictures of chimpanzees and humans (Figure 2-1-4a). Interestingly, we observed differential first fixation patterns between species when they were shown pictures of mammals. The chimpanzees were less likely than humans to locate their first fixation on the face when shown pictures of mammals (t-tests with Bonferroni correction; $t_{25} = 3.04$ $p = 0.005$; Figure 2-1-4a).

Figure 2-1-4

The high frequency of first fixations on the face was further examined by comparisons with subsequent fixations on this region (Figure 2-1-4b). Both chimpanzees and humans differentially fixated on the face across time when viewing pictures of mammals ($F_{4, 20} = 10.7$, $p < 0.001$; $F_{4, 80} = 56.3$, $p < 0.001$, respectively), chimpanzees ($F_{1.3, 6.6} = 6.10$, $p = 0.002$; $F_{4, 80} = 39.0$, $p < 0.001$, respectively), and humans ($F_{4, 20} = 10.6$, $p < 0.001$; $F_{2.4, 47} = 46.5$, $p < 0.001$, respectively; Figure 2-1-4b). Post-hoc comparisons (Dunnett tests with first fixation as a control) confirmed that the peak proportions of fixations on the face occurred with the same timing in both species: during second fixations for animal pictures and during first fixations for chimpanzee and human pictures.

Discussion

This study provides the first eye tracking data for chimpanzees viewing pictures. In comparing results for chimpanzees to those for humans, we discovered striking similarities in the gaze patterns of the two species, especially in terms of non-random patterns of scene viewing and the rapid detection of the face within the scene. We also observed several interesting differences between the species. Chimpanzees exhibited more rapid and broader shifts of the fixation location as well as shorter fixation durations in regard to the face region.

When shown pictures, the chimpanzees focused their fixations on informative regions, such as the body or face, in a manner very similar to humans (Figure 2-1-2 and 2-1-3), suggesting their active, voluntary control of gaze. These results were consistent with classic studies of humans (Buswell 1935; Yarbus 1967) and extend the observed paradigm to another species, the chimpanzee. Furthermore, both species repeatedly fixated on these informative regions, rather than scanning the entire area of the picture (Figure 2-1-2 and 2-1-4). In humans, this type of repetitive looking is known to indicate gaze control based on semantic informativeness of the region (Henderson & Hollingworth 1999a).

On the other hand, the two species also exhibited differences in their viewing patterns. The general pattern of chimpanzee eye movements was characterized by fixations that were greater in number, shorter in duration and more widely spread in space compared to those of humans (Figure 2-1-3). Two possible explanations for these interspecific differences involve the control and the function of eye movements. In terms of the mechanisms controlling eye movements, these differences may be influenced by the difference in high-level cognitive functioning between species. For example, the weaker inhibitory control of saccades in chimpanzees may have led to more frequent saccades and shorter fixations, similar to those observed in human infants and in some clinical patients (Hainline et al. 1984; Karatekin 2007). In addition, the shallower extraction of

information at each fixation point may have resulted in shorter fixation durations in chimpanzees, because fixations tend to remain in semantically informative regions longer, as has been shown in studies of humans (Henderson & Hollingworth 1999a). In regard to the functions of eye movements in the natural habitat, chimpanzees and humans may possess different strategies for processing scene information. Given that in humans, close or direct fixation is typically necessary to identify objects in scenes and to perceive their visual details (Henderson & Hollingworth 1999b), the more frequent and wider rotation of eyes that characterizes chimpanzees may represent a strategy for more quickly and widely retaining scene information.

The face region was most intensively viewed by both species compared to other parts of the body (Figure 2-1-2). This suggests that chimpanzees and humans share a common strategy of body viewing that is influenced by the informative quality of the face. Notably, longer durations of looking at the face occurred even when participants viewed pictures of mammals with various body/face morphologies, highlighting the generality of visual preferences for the face by both species. However, humans viewed the face longer compared to chimpanzees. This difference was attributed to longer fixation durations on the face in humans, because the number of fixations on this region did not differ between the species (Figure 2-1-3). The longer duration of fixations on the face in humans was explained in part by humans' longer fixation durations on the entire picture; however, there was a significant interaction between species and each feature of the pictures (Figure 2-1-3). In social interactions, long fixation on a face is perhaps a more intense signal of threat in non-human primates than in humans (Argyle 1967; Mendelson et al. 1982; Thomsen 1974). Thus, the observed pattern in chimpanzees - repetitive and brief looks at the face - may be explained by a higher likelihood of avoiding direct gaze contact. In addition, interestingly, chimpanzees looked at some parts of the body longer than humans did when viewing pictures of chimpanzees and humans (Figure 2-1-2). Furthermore, chimpanzees exhibited a larger number of fixations on other parts of the body

compared to humans (Figure 2-1-3). Together, these observations may also indicate that humans retrieve animal information primarily from the face, whereas chimpanzees gain relatively more information from the entire body.

The face region was detected at first sight by both species when they were shown pictures of chimpanzees and of humans (Figure 2-1-4a and b). This result echoes previous findings in humans (Fletcher-Watson et al. 2008; Hershler & Hochstein 2005) and chimpanzees (Tomonaga 2009), and indicates that rapid processing of person information also occurs in chimpanzees. In contrast, when shown pictures of mammals, the face was attended to at second rather than at first fixation by both species (Figure 2-1-4b), perhaps because both chimpanzees and humans responded less sensitively to the relatively unfamiliar body/face morphology of other mammals compared to the familiar morphology of their own or closely related species. Interestingly, when shown pictures of mammals, chimpanzees were less likely than humans to locate their first fixations on the face (Figure 2-1-4a). Because these chimpanzees have fewer opportunities than humans for observing non-primate mammals, these results further underscore the effects of prior experience on the rapid detection of faces. Our findings may be related to the “expertise effect” for face detection, which has previously been found in humans (Hershler & Hochstein 2005).

Further studies are necessary to address the hypotheses generated from our experiments. This study successfully provided a direct, quantifiable measurement of the eye movements of chimpanzees within a comparative cognitive context. What we could learn using this paradigm is promising; it guides an understanding of “what” details are informative and attractive for chimpanzees in their environments, as well as their serial processing of information, in other words, “where” and “when” their saccades go. Also, by presenting visual stimuli with other sensory cues such as their vocalizations, this paradigm can access how chimpanzees react to stimuli based on their preconception or motivation, which is analogous to task-dependent gaze behavior in humans (Yarbus

1967). In conjunction with recently developed non-invasive ERP measurement in chimpanzees (Ueno et al. 2008), it may open ways to examine when and how information is processed in the chimpanzee brain. Even for studies in more elusive concept such as theory of mind, the eye-tracking paradigm may provide a more objective measurement by compensating for the traditional looking time paradigm or utilizing, for example, an anticipatory looking paradigm (Southgate et al. 2007).

References

Argyle, M. 1967 Non-verbal communication in human social interaction. In Nonverbal communication (ed. A. Mehrabia), pp. 243-269. Chicago: Aldine-Atherton

Buswell, G. T. 1935 How People Look at Pictures: A Study of the Psychology of Perception in Art: The University of Chicago press.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R. & Benson, V. 2008 Rapid detection of person information in a naturalistic scene. *Perception* 37, 571-583.

Fujita, K. 1993 Role of some physical characteristics in species recognition by pigtail monkeys. *Primates* 34, 133-140.

Fujita, K. & Matsuzawa, T. 1986 A new procedure to study the perceptual world of animals with sensory reinforcement: Recognition of humans by a chimpanzee. *Primates* 27, 283-291.

Gothard, K. M., Erickson, C. A. & Amaral, D. G. 2004 How do rhesus monkeys (Macaca mulatta) scan faces in a visual paired comparison task? *Anim Cogn* 7, 25-36.

Hainline, L., Turkel, J., Abramov, I., Lemerise, E. & Harris, C. M. 1984 Characteristics of saccades in human infants. *Vision Res* 24, 1771-80.

Henderson, J. M. & Hollingworth, A. 1999a High-level scene perception. *Annu Rev Psychol* 50, 243-271.

Henderson, J. M. & Hollingworth, A. 1999b The role of fixation position in detecting scene changes across saccades. *Psychol Sci* 10, 438-443.

Hershler, O. & Hochstein, S. 2005 At first sight: A high-level pop out effect for faces. *Vision Res* 45, 1707-1724.

Karatekin, C. 2007 Eye tracking studies of normative and atypical development. *Dev Rev* 27, 283-348.

Keating, C. F. & Keating, E. G. 1982 Visual scan patterns of rhesus monkeys viewing faces.

Perception 11, 211-219.

Kobayashi, H. & Kohshima, S. 2001 Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J Hum Evol* 40, 419-435.

Matsuzawa, T. 2006 Sociocognitive development in chimpanzees: A synthesis of laboratory work and fieldwork. In Cognitive Development in Chimpanzees (ed. T. Matsuzawa, M. Tomonaga & M. Tanaka), pp. 3-33. Tokyo: Springer.

Matsuzawa, T., Tomonaga, M. & Tanaka, M. 2006 Cognitive development in chimpanzees. Tokyo: Springer.

Mendelson, M. J., Haith, M. M. & Goldmanrakic, P. S. 1982 Face Scanning and Responsiveness to Social Cues in Infant Rhesus-Monkeys. *Dev Psychol* 18, 222-228.

Nahm, F. K. D., Perret, A., Amaral, D. G. & Albright, T. D. 1997 How do monkeys look at faces? *J Cognitive Neurosci* 9, 611-623.

Southgate, V., Senju, A. & Csibra, G. 2007 Action anticipation through attribution of false belief by 2-year-olds. *Psychol Sci* 18, 587-592.

Tanaka, M. 2003 Visual preference by chimpanzees (*Pan troglodytes*) for photos of primates measured by a free choice-order task: implication for influence of social experience. *Primates* 44, 157-165.

Thomsen, C. E. 1974 Eye contact by non-human primates toward a human observer. *Anim Behav* 22, 208.

Tomonaga, M. 1994 How laboratory-raised Japanese monkeys (Macaca fuscata) perceive rotated photographs of monkeys: Evidence for an inversion effect in face perception. *Primates* 35, 155-165.

Tomonaga, M. (2010). Chimpanzee eyes have it? Social cognition on the basis of gaze and attention

from the comparative-cognitive-developmental perspective. In E. Lornsdorf, S. Ross & T. Matsuzawa (Eds.), *The mind of the chimpanzee: Ecological and empirical perspectives.* Chicago: University of Chicago Press.

Ueno, A., Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., Matsuda, G., Fukushima, H., Hiraki, K., Tomonaga, M. & Hasegawa, T. 2008 Auditory ERPs to Stimulus Deviance in an Awake Chimpanzee (*Pan troglodytes*): Towards Hominid Cognitive Neurosciences. *PLoS ONE* 3.

Van Hooff, J. 1967 The Facial Displays of the Catanhine Monkeys and Apes. In Primate Ethology (ed. D. Morris), pp. 7-68: Aldine Transaction.

Yarbus, A. L. 1967 Eye movements during perception of complex objects. In Eye Movements and Vision, pp. 171-211. New York: Plenum Press.

Figure captions

Figure 2-1-1: An example of (a) non-primate mammal, (b) chimpanzee, and (c) human picture presented in this study. Typical scanpaths of a chimpanzee (yellow) and a human (red) are superimposed on each picture. Each circle represents a fixation, which is linked to the adjacent fixation by a line. A longer fixation is drawn as a larger circle.

Figure 2-1-2: Gaze time (the sum of fixation durations; ms + SEM) to each scene/body feature of animal pictures by both species (6 chimpanzees and 21 humans). The schematic drawings are examples of the defined features used for quantitative analysis and indicate the relative size of each. The mean scores over 24 trials are shown. Each picture was presented for 3 s.

Figure 2-1-3: General differences in eye movements by both species, indicated by (a) number of fixations (+ SEM), (b) fixation duration (ms + SEM) and (c) distances between fixations (degree + SEM) on each scene/body feature.

Figure 2-1-4: (a) The proportion of first fixations (+ SEM) on each scene/body features by both species, averaged over 24 trials. (b) The proportion of fixations (± SEM) on the face

Figures

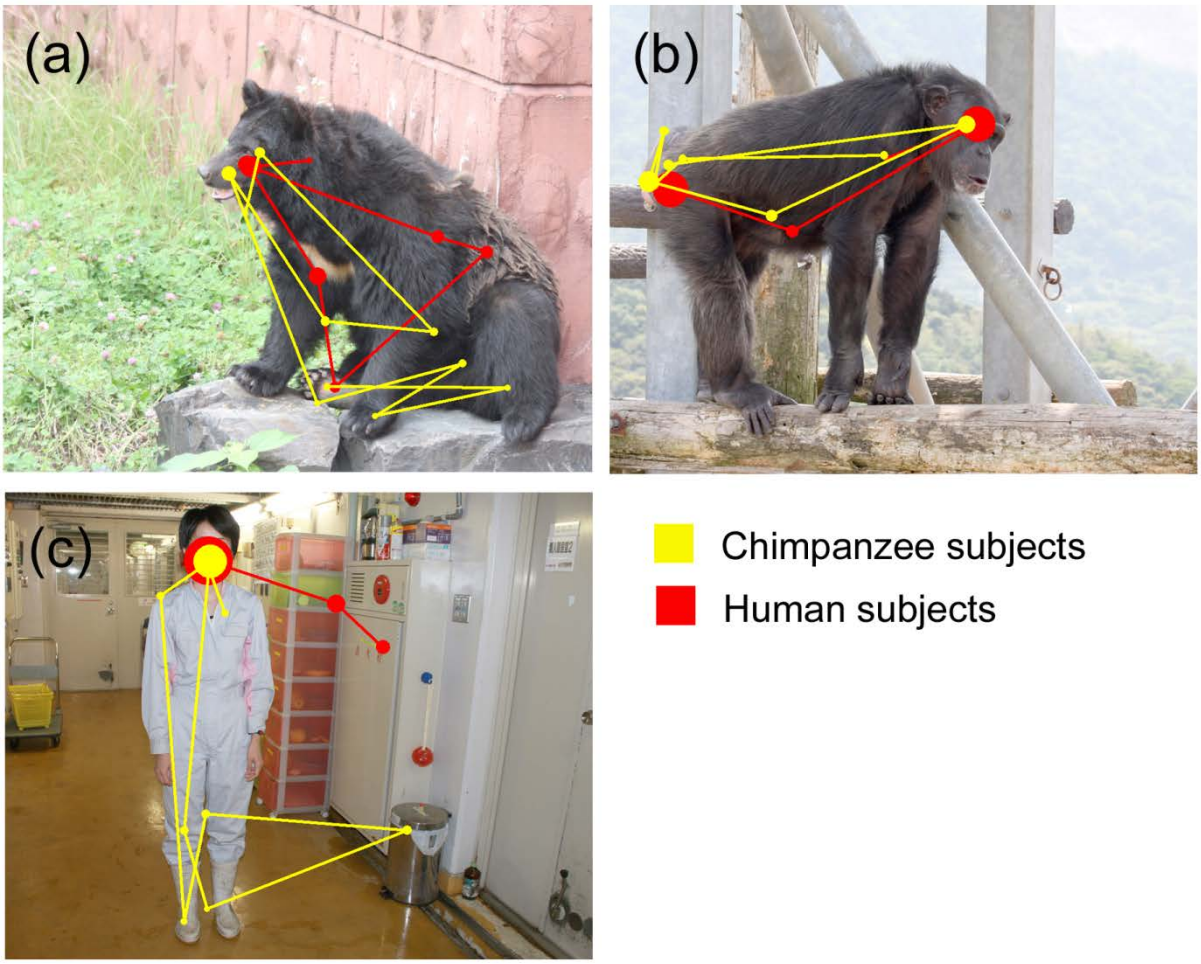


Figure 2-1-1

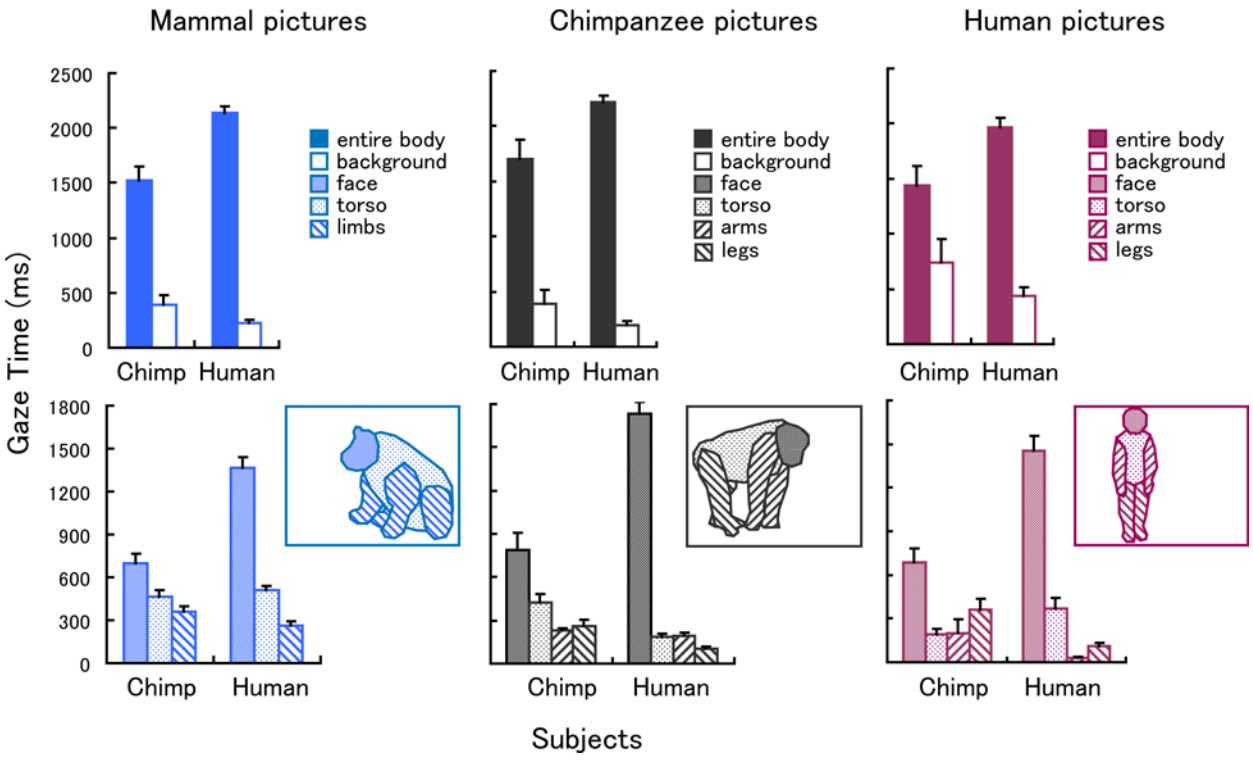


Figure 2-1-2

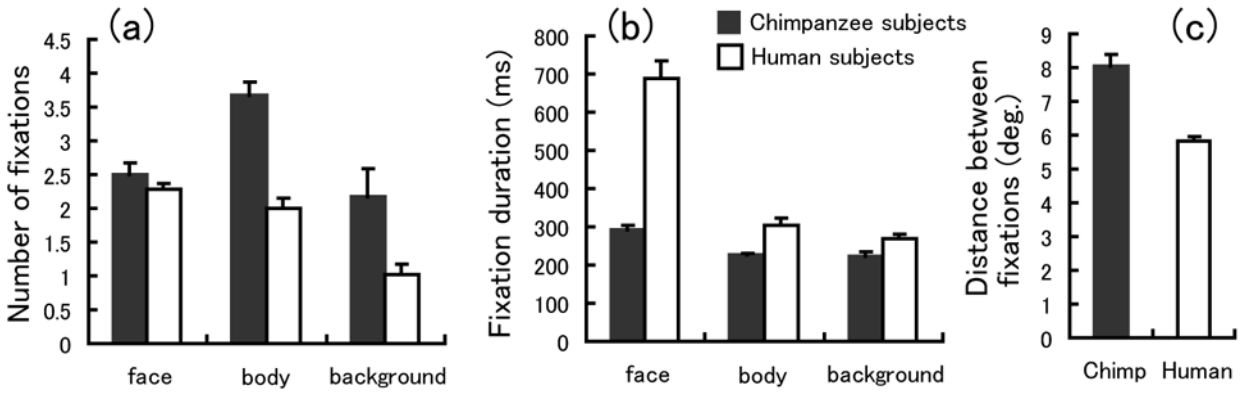


Figure 2-1-3

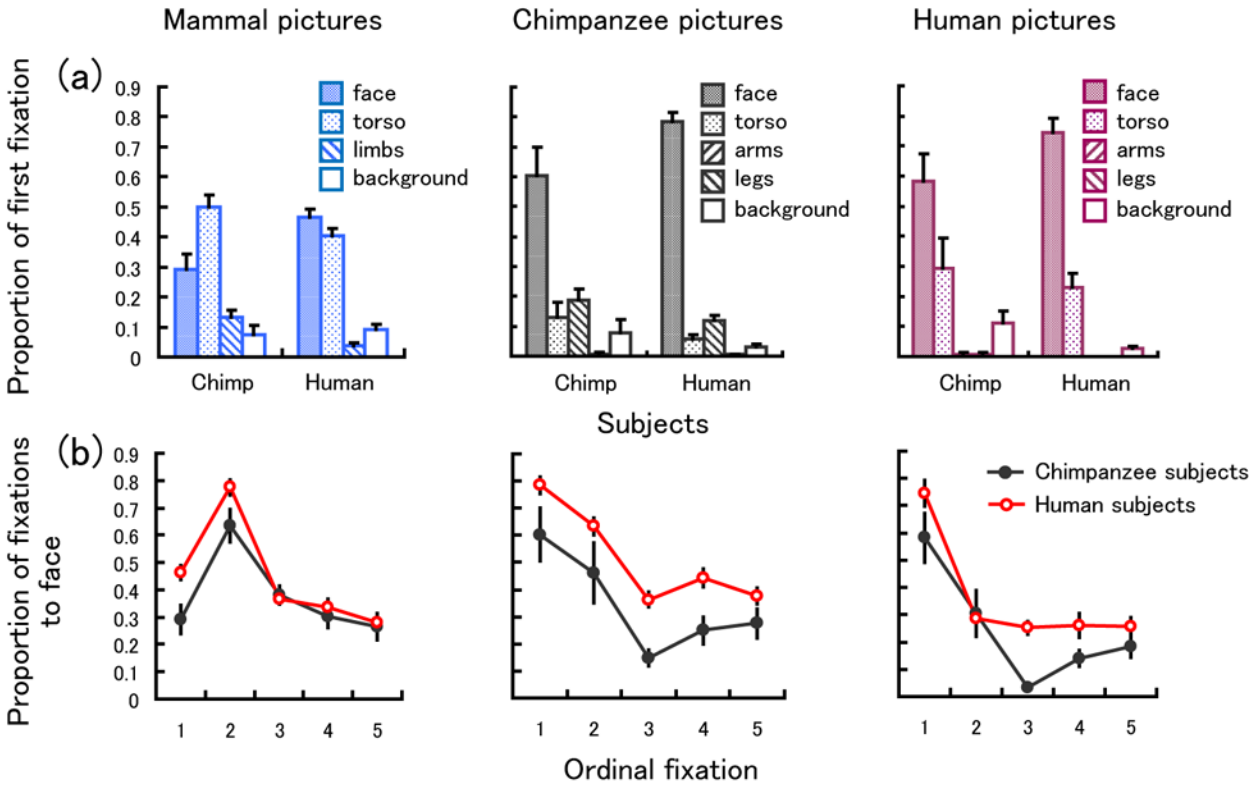

Mammal pictures
Chimpanzee pictures
Human pictures
(a)
Proportion of first fixation
face
torso
limbs
arms
legs
background
Chimp
Human
(b)
Proportion of fixations to face
Subjects
Chimpanzee subjects
Human subjects
Ordinal fixation

Figure 2-1-4

Chapter 2-2.
Perceptual mechanism underlying gaze guidance in chimpanzees and humans

Introduction

Eye-tracking methodology in human and nonhuman primates has been used for over 50 years (Fuchs, 1967; Yarbus, 1967). Eye-movement patterns of nonhuman primates show a significant degree of similarity with those of humans under similar experimental conditions (Kano and Tomonaga, 2009; Shepherd, Steckenfinger, Hasson, and Ghazanfar, 2010). Comparative studies of human and nonhuman primates have directly compared the species in order to clarify both similarities and differences in their eye-movement characteristics (Dahl, Wallraven, Bulthoff, and Logothetis, 2009; Gothard, Brooks, and Peterson, 2009; Guo, Robertson, Mahmoodi, Tadmor, and Young, 2003; Kano and Tomonaga, 2009; Keating and Keating, 1982; Nahm, Perret, Amaral, and Albright, 1997; Shepherd et al., 2010). Those similarities and differences have been an important source of information for the study of the evolution of visual behavior, social perception, and high-level cognition (Kano and Tomonaga, 2009; Shepherd et al., 2010). Although apes have been essential for this comparative approach, their eye-movement characteristics are largely unknown compared to those of the well-studied macaque species.

Recently, eye-tracking studies in chimpanzees (*Pan troglodytes*), the species most closely related to humans, have been reported (Hattori, Kano, and Tomonaga, 2010; Hirata, Fuwa, Sugama, Kusunoki, and Fujita, 2010; Kano and Tomonaga, 2009, 2010). Those studies have presented naturalistic images of scenes (including faces, bodies, etc.) to chimpanzees and humans under free-viewing conditions and compared their fixation patterns under similar experimental conditions. There are several advantages of comparing chimpanzees and humans for a free-viewing task. First, chimpanzees are the species most closely related to humans and are known to have similar perceptual mechanisms (Matsuno, Kawai, and Matsuzawa, 2004; Matsuzawa, 1985, 1990; Tomonaga and Matsuzawa, 1992). Second, the demands of a free-viewing task are small; the participants of both species need only look at the stimuli spontaneously and are not trained to solve particular problems using their eye movements. Third, for the same reason, a free-viewing task is relatively independent of the effect of reward or training. Therefore, we are able to efficiently and directly compare the species, find both similarities and differences between them, and discuss the

extent to which chimpanzees and humans are similar and different in their perception and cognition.

In the previous study comparing chimpanzees and humans in a free-viewing task, it was found that the species were very similar in terms of where to fixate (i.e. scanpath similarity). For example, when presented with a scene including an entire body of a chimpanzee, a human, or another animal, both chimpanzees and humans concentrated fixations on the body, especially the face, rather than on the background. In addition, both species fixated on the face immediately after the image presentation (within the first few fixations). However, those responses differed quantitatively between the species; humans showed a higher proportion of face fixations than did chimpanzees. There seem to be several functional reasons for those similarities and differences between the species. First, faces are the most important source of social information (such as individuality and emotions) for both chimpanzees and humans (Chevalier-Skolnikoff, 1973; Parr, Dove, and Hopkins, 1998), and thus frequent inspection and immediate detection of facial characteristics may benefit them by enabling them to obtain such information efficiently. Second, humans have a specific form of facial communication; humans often engage in lengthy face-to-face communication, accompanied by intense eye contact (Argyle and Cook, 1976). Therefore, more frequent inspection of faces may benefit humans more specifically than chimpanzees in the context of their own form of facial communication.

There seem to be several factors that determine such similarities and differences. These include, for example, the perception of low-level visual properties (e.g. color, form), the perception of bodies and faces, and knowledge about the scenes (which the viewers had obtained through daily lives or experimental instructions). Previous studies using forced-choice discrimination paradigms have found that the perceptions of low-level visual properties involving color (Matsuno et al., 2004; Matsuzawa, 1985), form (Matsuzawa, 1990; Tomonaga and Matsuzawa, 1992) are largely similar between chimpanzees and humans. In addition, the mechanisms involving advanced social perceptions involving faces (Parr et al., 1998; Parr, Hecht, Barks, Preuss, and Votaw, 2009; Tomonaga, 2007, 2010; Tomonaga and Imura, 2009) and bodies (Tomonaga and Imura, 2008) are also similar between the species.

Because of these similarities between chimpanzees and humans, it is expected that the influence of low-level stimulus properties on their eye-movement patterns appears similarly in the two species and does not play a critical role in explaining for the overall similarities and differences between the species. However, there are no quantitative data to support this assumption. It is important to separate low-level from higher-level influences on eye-movement patterns in order to provide a foundation for direct comparison between the species. Therefore, this study aimed to elucidate the influence of low-level stimulus properties on the eye-movements of chimpanzees and humans.

We used two approaches in order to separate low-level from high-level influences on eye movements. First, to simulate responses to local stimulus properties, we used the well-established bottom-up saliency model (Itti and Koch, 2000; Walther and Koch, 2006). This model estimates the local saliency of an image based on its low-level components -- such as color, intensity, and component orientations -- and predicts the locations of attention based on these local saliency values. The second approach used global manipulation of stimulus properties (e.g., stimulus properties such as color, configuration, frequency components, orientation, complexity, and location of scene features) and observed how participants changed their patterns of scanning in response to the manipulations.

In this study, we used similar stimulus sets to those used by Kano and Tomonaga (2009) and analyzed the participants' responses to social stimuli, especially to the face, as a main measure. In the previous study, chimpanzees and humans fixated the face more frequently than any other part of the scene. The frequent fixation to the face is most likely caused not only by the low-level saliency of the faces, but also by the participants' sensitivity to the social stimuli. This study aimed to investigate the extent to which such facial fixation patterns could be explained by the bottom-up saliency model and could be influenced by the global manipulation of stimulus properties in the scene.

Therefore, the topics we addressed in this study were as follows. (1) The degree of similarity in fixation distribution patterns between chimpanzees, humans, and those predicted by the

bottom-up saliency model; we expected similar patterns of fixation distribution between chimpanzees and humans even when we controlled for the low-level saliency. (2) The extent to which the gaze of chimpanzees and humans is attracted by low-level saliency. (3) The extent to which the two species' facial fixation is influenced by the global manipulation of stimulus properties in the scene. For (2) and (3), again we expected a significant degree of similarity between the species given their perceptual similarities.

Method

Participants

Six chimpanzees (five females, one male; aged 9–31 years) and 16 humans (11 females, five males; aged 18–31 years; all Japanese) participated in this experiment. The chimpanzees were members of a social group of 14 individuals living in enriched outdoor compounds and attached indoor residences (Matsuzawa, Tomonaga, and Tanaka, 2006). They were highly experienced in observing pictorial representations appearing on a computer screen (Matsuzawa et al., 2006). No food or water deprivation occurred during the study period. Care and use of the chimpanzees adhered to the 2002 version of the Guidelines for the Care and Use of Laboratory Primates published by the Primate Research Institute, Kyoto University. The experimental protocol was approved by the Animal Welfare and Care Committee of the Institute and by the Animal Research Committee of Kyoto University. The human participants were graduate and undergraduate students, who participated in the experiment voluntarily. Informed consent was obtained from all human participants.

Apparatus

Both species used the same apparatus, in order to ensure the possibility of direct comparison between the species. Participants sat still and unrestrained in an experimental booth, with the eye-tracking apparatus and the experimenter separated by transparent acrylic panels. A table-mounted eye tracker measured their eye movements using infrared corneal reflection techniques (60 Hz; Tobii X120, Tobii Technology AB). This eye-tracker has wide-angle lenses (±40 degrees in the semicircle above the camera) and thus obviated the necessity to restrain the subjects. The eye-tracker and the 17-inch LCD monitor (1280×1024) were mounted on a movable platform, and the distance between the platform and the participants was adjusted to the point at which the gaze was most accurately recorded (60 ±10 cm). This flexible adjustment of the distance between the platform and the participants enabled us to record the gaze movements of chimpanzees without any head restraint. The participant's gaze was recorded as a relative coordinate with respect to the monitor size (i.e. not as the gaze angle). One degree of gaze angle corresponded to approximately 1

cm on the screen at a typical 60-cm viewing distance.

As a result of the training conducted during the study performed by Kano and Tomonaga (2009), the chimpanzees were already skilled at sitting still in front of an eye-tracker and looking upon request at a fixation point that appeared on the screen. Five-point calibration was conducted for humans; for chimpanzees, the calibration points were reduced to two in order to decrease the time required for each calibration process. However, for chimpanzees, the calibration was repeated until the maximum accuracy was obtained. The accuracy was checked by presenting to both species five fixation points on the screen. Using these calibration procedures, six participants of both species were tested for accuracy, and the errors were found to be small and comparable between the species (mean errors of 0.62 ±0.06 and 0.52 ±0.05 cm ± s.e.m. on the monitor for chimpanzees and humans, respectively). The drift (the calibration error due to changes occurring in the eye surface) was checked occasionally by presenting the fixation points to the participants again.

Stimuli

We prepared 20 color photographs of naturalistic scenes containing a human figure (Figure 2-2-1). We used only human figures (all Japanese; no chimpanzees or other animals) in this study because a previous study using an identical experimental procedure (Kano and Tomonaga 2009) found similar fixation patterns in both species for all animal figures. These 20 images served as the control condition. Eight experimental conditions were additionally prepared (for the details of manipulation procedure, see Table 2-2-1). In the monochrome, line drawing, and schematic drawing conditions, we eliminated color, low-spatial frequency component, and complex lines, respectively, from the original color scene and aimed to examine the influence of realistic appearance of a scene on the participants' response to the faces. In the blurred and silhouette conditions, we blurred and eliminated local features of face and body from the scene and aimed to examine the influence of those features on the responses. In the upside down and scramble conditions, we inverted and scrambled the scene, respectively, and aimed to examine the influence of orientation and arrangements of bodily parts on the response (i.e. we checked whether participants used only information that the head is above the body). In the headless condition, we eliminated the head from

the body and aimed to examine whether participants used only bodily information to fixate the location where the head ought to be. Overall, these conditions aimed to observe whether participants used multiple cues to detect the location of faces in the scene. Each experimental condition was represented by five examples created by manipulating the control images. These five examples were pseudo-randomly selected from the entire set of 20 control images so that each control image was used at least once in the experimental conditions. In total, 60 stimuli were used (40 experimental and 20 control images). The images were converted into 1000 × 800 pixel images with surrounding gray frames (1280 × 1024 pixels in total; 37 × 30 degrees at a typical 60-cm viewing distance). We used Adobe Photoshop CS3 to process the images.

Figure 2-2-1

Procedure

Procedural differences for testing chimpanzees and humans were minimized to allow for direct comparisons between species. In each trial, an image was presented after participants focused on a fixation point that appeared at a random position on the screen. Participants were then allowed to view images freely. The participants of both species rarely kept gazing at the fixation point after the image presentation (i.e. spontaneous scanning was almost always observed). Stimuli were presented for 3 sec each. The presentation order of conditions and trials were randomized for each participant so that the same conditions were not presented more than three times in succession. 20 other stimuli depicting various interesting scenes (e.g. pictures of funny faces) were presented occasionally during the sessions in order to keep the participants interested. The entire session therefore consisted of 80 trials: 60 experimental stimuli and 20 others. The entire session was conducted on a single day for humans, whereas the session was divided among 15 days for the chimpanzees in order to decrease the time required for each daily experiment (each day used six examples for the chimpanzees). In a preliminary session, we confirmed that our human participants showed similar scanning patterns of bodies/faces when tested on separate days (comparing the

results from this study with those from Kano and Tomonaga (2009)). Daily experiments lasted 10–15 min for the chimpanzees and 20 min for the humans. Human participants received book coupons as rewards after the session, and chimpanzees received a small piece of apple after each trial. The reward was given for chimpanzees in return for the initial fixation at the beginning of the trial, and thus was given independently of their viewing behavior during the image presentation. Overall, those procedural differences between the species were made in an effort to increase the motivation of both species to participate in the daily experiment, and to keep their interest during the presentation of each image (3 sec). Trials in which participants only glanced at the monitor (one or two fixations) were repeated after the whole session and were replaced by the new trials. As a result, we had no loss of trials for both species.

Data analysis

Fixation definition. A fixation was scored if the gaze remained stationary within a radius of 50 pixels for at least 75 ms (more than five measurement samples). Otherwise, the recorded sample was defined as part of a saccade. The records during the first 200 ms were eliminated from the analysis, thereby eliminating fixations that followed the offset of the initial fixation point.

Area of Interest (AOI). Each stimulus was divided into areas of interest (AOI) for the purpose of quantitative comparison. Each scene was divided into background, face, torso, arms, and legs. Each AOI was drawn 20 pixels larger than the precise outline of the features to avoid errors in gaze estimations. The AOI of background, torso, legs, arms, and face were laid above in this order (i.e. face is the topmost). If two or more AOIs were duplicated, the samples were added to the upper AOI.

Chance level. The chance level was set on the assumption that participants would view images randomly. However, the participants generally showed a central bias in fixation distribution, while the model did not (evident by inspection of Figure 2-2-2). This needs to be controlled to compare participants with the model, because such central bias is known to be caused either by the participants' bias in scanning images or by the experimenter's bias in the arrangement of main objects in the scene (Henderson, Brockmole, Castelhano, and Mack, 2007; Tatler, 2007) (i.e., caused

independently of the low-level stimulus properties). Therefore, in this study, we modified the definition of chance level by controlling for such particular bias shown by each participant. Specifically, we compared the particular scanpath, which was obtained from a participant (or the model) in a trial, with all the other scanpaths, which were obtained from the same participants (or the model) in all the other trials of the experiment. All data shown in this study were calculated as differences between the value obtained from the particular scanpath and the mean value obtained from the other control scanpaths (i.e., the chance level).

Figure 2-2-2

Saliency model. We used the well-established bottom-up model to estimate the low-level saliency of the images (Itti and Koch, 2000; Walther and Koch, 2006). This model processes the image with respect to several features -- such as color (red-green, blue-yellow), intensity, and orientation (0, 45, 90, 135 degrees) -- then extracts the local discontinuities in each feature, and finally combines them into a single 'saliency map' (Figure 2-2-1). The model then predicts a scanpath based on the saliency map, selecting salient locations in order of decreasing saliency. In this experiment, the saliency maps and the model scanpaths were generated by Saliency Toolbox 2.2 (http://www.saliencytoolbox.net) in Matlab with all-default parameters. We used the original resolution of images (1280×1024; including the surrounding gray frame) for the simulation in the model. Because this model does not predict the duration of each fixation, we arbitrarily set the scanpath length of the model as 9 fixations (about as long as chimpanzee scanpaths in 3-s viewing) to compare the model with the participants. There is no variance in the output when repeating the simulation.

To determine the saliency value at each fixated area, we employed the following procedures. First, saliency value was normalized within each map to a range of 0 (not salient) to 1 (highly salient). Second, to avoid errors in gaze estimation, the saliency map was divided into a 12×9 grid, and all saliency values (i.e. 1280×1024 samples, in total) were summed within each grid (i.e. each grid had approx. 100×100 samples). The fixated area was defined as the grid where the fixation

was observed, and the saliency value of each grid was used for the saliency value at each fixated area.

Results

Figure 2-2-3a shows the distribution patterns of fixation over the scene in each species/model (the data were sampled from 20 control images). Comparing between species and between AOIs, we found a significant interaction ($F_{2.3, 47} = 9.52$, $p < 0.001$, $\eta^2 = 0.32$)[1] because chimpanzees distributed their fixations over the scene more widely than did humans. Comparing between AOIs respectively for each species, we found significant main effects for both chimpanzees ($F_{1.8, 9.3} = 23.80$, $p < 0.001$, $\eta^2 = 0.82$) and humans ($F_{2.4, 37} = 358.86$, $p < 0.001$, $\eta^2 = 0.96$) because both species showed higher proportion of fixations on particular areas (the bodies, especially faces, rather than backgrounds) than would be expected by chance (represented as zero in the figures). This pattern of results emerged even when the model was subtracted from each species: chimpanzees ($F_{1.8, 9.3} = 5.61$, $p = 0.003$, $\eta^2 = 0.52$) and humans ($F_{2.4, 37} = 159.74$, $p < 0.001$, $\eta^2 = 0.91$). This pattern emerged no later than the first two fixations, as shown in Figures 2-2-3b and 2-2-3c, and is consistent with the previous reports in humans (Crouzet, Kirchner, and Thorpe, 2010; Fletcher-Watson, Findley, Leekam, and Benson, 2008; Honey, Kirchner, VanRullen, 2008). The global similarities in distribution patterns of fixation among chimpanzees, humans, and the model suggest that the saliency model partially (but not fully) explained those patterns for the two species. Although chimpanzees were more similar to the model than were humans in that regard, it should be noted that this does not mean that the low-level visual saliency influenced chimpanzees more strongly than humans; this means that chimpanzees distributed their fixations over the scene more widely than did humans, but less widely than did the model.

Figure 2-2-3

Indeed, chimpanzees and humans did not significantly differ in their responses to low-level visual saliency. There was no significant effect of species in the saliency values at fixation (Figure 2-2-4); neither the main effect of species ($F_{1, 20} = 0.014$, $p = 0.90$, $\eta^2 = 0.001$) nor the interaction between species and fixation order ($F_{5, 100} = 0.46$, $p = 0.80$, $\eta^2 = 0.023$) was significant.

Overall, however, both species fixated on salient regions in the scene more than would be expected by chance: the mean saliency values for the first 6 fixations were significantly higher than zero in both chimpanzees ($t(5) = 9.83$, $p < 0.001$) and humans ($t(15) = 19.27$, $p < 0.001$). This pattern emerged more strongly for the earlier than for the later fixations: saliency value decreased as a function of increasing fixation order ($F_{5, 100} = 3.20$, $p = 0.010$, $\eta^2 = 0.13$). These results suggest that the saliency model predicted the distribution patterns of fixation in both chimpanzees and humans better than chance, especially for the early fixations. However, it should be noted that this result does not necessarily mean that the low-level saliency alone influenced the species' distribution patterns of fixation, because such frequently fixated areas (e.g., bodies and faces) were in general more visually salient (because of the complexity of lines, for example) as well as more informative than the other areas of the scene (Figure 2-2-3; refer to (Henderson et al., 2007) for a similar discussion).

Figure 2-2-4

We then examined the effect of image manipulations on the fixation patterns of chimpanzees and humans (Figure 2-2-5). Figure 2-2-5b shows the proportion of fixations on the faces as a function of image manipulations. There was no interaction of species with condition ($F_{3.5, 70} = 1.13$, $p = 0.34$, $\eta^2 = 0.05$). The main effect of species was significant: humans showed a higher proportion of fixations on faces than did chimpanzees ($F_{1, 20} = 5.51$, $p = 0.029$, $\eta^2 = 0.21$), which is consistent with the aforementioned result. The main effect of condition ($F_{3.5, 70} = 5.33$, $p < 0.001$, $\eta^2 = 0.21$) was significant: participants showed a lower proportion of face fixations in the headless than the other conditions (as was revealed by the pair-wise comparisons with Bonferroni's correction).

Figure 2-2-5

However, even in the headless condition, both species showed a higher proportion of fixations on the face original locations than would be expected by the model (as was revealed by the

post-hoc *t*-tests). This means that even when a head was actually absent from the scene, both species concentrated fixations on the area where the face would have been (i.e. above the body).

Figure 2-2-5c shows the mean saliency values at the first 6 fixations as a function of image manipulations. The main effect of condition was significant ($F_{8, 160} = 46.93$, $p < 0.001$, $\eta^2 = 0.70$), probably modulated by saliency (or informativeness) in local features of the scene, which was an outcome of image manipulations. Importantly, there was no effect of species despite these image manipulations, either the main effect of species ($F_{1, 20} = 0.017$, $p = 0.89$, $\eta^2 = 0.001$) or the interaction between species and condition ($F_{8, 160} = 1.18$, $p = 0.31$, $\eta^2 = 0.05$).

Discussion

Chimpanzees and humans distributed fixations over the scene non-randomly, and showed higher fixation proportions on particular areas of the scene, especially faces, than would be expected by the saliency model. However, humans showed an even higher proportion of fixation on the bodies and faces than did chimpanzees. These results emerged even at the first two fixations, at the earliest moments of scene inspection, suggesting that those fixation patterns reflect automatic rather than voluntary control of gaze. Saliency values of chimpanzees and humans in the fixated region were higher than would be expected by chance, suggesting that low-level saliency partially (but not fully) predicted the species' distribution patterns of fixation. However, chimpanzees and humans did not significantly differ in their responses to low-level saliency. None of global manipulations of stimulus properties in the scene (color, configuration, frequency components, orientation, complexity, and local features) critically altered both species' strong tendency toward fixating faces, suggesting that both species used multiple cues to fixate faces. In addition, although those manipulations changed the extent to which low-level saliency influenced both species, chimpanzees and humans did not differ in the degree of change in the response.

Therefore, chimpanzees and humans seem to be qualitatively similar in the sense that both species have an enhanced perceptual mechanism to guide their fixation location, one which is more complex than would be presumed on the basis of the saliency model (i.e. color, intensity, and orientations), and have multiple strategies to perceive the location of faces. Quantitatively, these two species did not differ significantly in their responses to low-level saliency, suggesting that they have similar perceptual mechanisms to guide the fixation locations.

Einhäuser et al. (Einhuäuser, Kruse, Hoffmann, and König, 2006) used the standard saliency model to predict the fixation location of monkeys (rhesus macaques) and humans when presented with the still images of naturalistic scene (without social contents). They found that monkeys and humans did not differ significantly in their responses to low-level saliency when viewing those images, which is consistent with the present study comparing chimpanzees and humans. However, when the luminance-contrast (or the saliency) was manipulated locally in the

image, the monkeys responded to those manipulated areas more strongly than did the humans. In the similar analysis to that of Einhäuser et al. (2006) and this study, Berg et al. (Berg, Boehnke, Marino, and Munoz, and Itti, 2009) found that, when presented with dynamic scenes including various social, non-social, and narrative contents, humans responded to the low-level visual saliency more strongly than did monkeys (perhaps because monkeys tended to move their eyes independently of the stimuli (e.g. inattentiveness to the stimuli) or show a large degree of individual differences in their fixation patterns), which is somewhat inconsistent with Einhäuser et al. (2006) and this study. Therefore, multiple factors seem to be involved in the species difference in the responses to the low-level saliency. To clarify those factors, it is necessary to directly compare between the three species for their fixation patterns when presented with various contents of still and dynamic scenes.

Cerf et al. (Cerf, Harel, Einhäuser, and Koch, 2008) have shown that the addition of a "face channel" into the standard saliency model better predicts the fixation patterns of human participants viewing a naturalistic scene that includes faces. They used an established face detector algorithm for that purpose, which predicts the location of faces based on local facial features (e.g. local discontinuities in intensity around eye and nose regions). The distribution patterns of fixation observed in this study suggest that chimpanzees and humans have such a face perception channel in addition to the low-level channels. However, the mechanism underlying such a face channel seems more complex in chimpanzees and humans than would be assumed by the face detector algorithm. This is because chimpanzees and humans concentrated fixations on the faces even when local features of faces were significantly reduced (schematic and blurred) or completely silhouetted out of the scene. They did so even when the faces were removed completely (headless), suggesting that chimpanzees and humans can use the bodily configuration alone to fixate where faces ought to be. On the other hand, chimpanzees and humans also seem to be able to use local cues to fixate faces, because they concentrated fixations on the face parts even when bodily configuration was disrupted (scrambled). Therefore, chimpanzees and humans seem to have an enhanced perceptual mechanism to guide their fixations to a face, a mechanism that is more complex than would be assumed by the standard saliency model or the saliency model combined with face detection.

Notwithstanding those similarities between the species, chimpanzees and humans differ quantitatively in the distribution patterns of fixations. Humans showed a higher proportion of fixations on bodies and faces than did chimpanzees. As clarified above, it is unlikely that this species difference resulted from their differential responses to the low-level visual properties (or in their differential tendencies for central bias). It is also unlikely that this species difference resulted from the use of human, as opposed to chimpanzee figures as stimuli, because a previous study (Kano and Tomonaga, 2009) obtained the same patterns of results when using chimpanzees and other mammals as the stimulus models. Therefore, we interpreted the results in the following two ways. First, although the results suggested that both species have similar mechanism to guide their gaze to the social stimuli (body and face), those mechanisms may operate differently in each species. For example, humans may put more emphasis on the body/face channels to create the saliency map, and so humans may perceive bodies and faces as more salient than chimpanzees do. Second, humans, compared to chimpanzees, may have a stronger tendency to process scenes in a top-down rather than a bottom-up manner, and thus would be expected to show a higher proportion of fixations on the semantically informative areas such as bodies and faces. Further studies are necessary to test these two possibilities.

In summary, this study presented, to chimpanzees and humans, naturalistic (unmanipulated) scenes including body, face, and their manipulated representations. We then compared among the two species and the saliency model for the fixation patterns on the images. We found that the saliency model did not fully predict the fixation patterns actually observed in chimpanzees and humans. In addition, both species used multiple cues to fixate the face. There was no evidence suggesting any differences between chimpanzees and humans in the perception of low-level saliency (e.g. color, intensity, or orientations). Therefore, we showed a substantial amount of similarities in the perceptual mechanism underlying gaze guidance between chimpanzees and humans, and thereby offer a foundation for the direct comparison between the species. Further studies are necessary to elucidate the high-level similarities and differences between the species (e.g. social sensitivity, knowledge-based attention).

References

Argyle, M., Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *J Vis, 9*(5), 1-15.

Cerf, M., Harel, J., Einhuäuser, W., Koch, C. (2008). Predicting human gaze using low-level saliency combined with face detection. *Adv Neural Inform Process Syst, 20*, 241-248.

Chevalier-Skolnikoff, S. (1973). Facial expression of emotion in nonhuman primates. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review* (pp. 11-89). New York: Academic Press.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. Fast saccades toward faces: Face detection in just 100 ms. *J Vis, 10*(4), 1-17.

Einhäuser, W., Kruse, W., Hoffmann, K. P., & König, P. (2006). Differences of monkey and human overt attention under natural conditions. *Vision Res, 46*(8-9), 1194-1209.

Dahl, C. D., Wallraven, C., Bulthoff, H. H., Logothetis, N. K. (2009). Humans and macaques employ similar face-processing strategies. *Curr Biol, 19*(6), 509-513.

Fuchs, A. F. (1967). Saccadic and smooth pursuit eye movements in the monkey. *J Physiol, 191*(3), 609-631.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37*, 571-583.

Gothard, K. M., Brooks, K. N., Peterson, M. A. (2009). Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim Cogn, 12*(1), 155-167.

Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y., Young, M. P. (2003). How do monkeys view faces?—a study of eye movements. *Exp Brain Res, 150*(3), 363-374.

Hattori, Y., Kano, F., Tomonaga, M. (2010). Differential sensitivity to conspecific and allospecific

cues in chimpanzees and humans: A comparative eye-tracking study. *Biology Lett. 6*(5), 610-613

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 537-562). Neitherlands: Elsevier.

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *J Vis, 8*(12), 1-13.

Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., Fujita, S. (2010). Facial perception of conspecifics: chimpanzees (*Pan troglodytes*) preferentially attend to proper orientation and open eyes. *Anim Cogn, 13*(5), 679-688.

Itti, L., Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res, 40*(10-12), 1489-1506.

Kano, F., Tomonaga, M. (2009). How chimpanzees look at pictures: a comparative eye-tracking study. *Proc Roy Soc B, 276*(1664), 1949-1955.

Kano, F., Tomonaga, M. (2010). Face scanning in chimpanzees and humans: continuity and discontinuity. *Anim Behav, 79*, 227-235.

Keating, C. F., Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception, 11*(2), 211-219.

Matsuno, T., Kawai, N., Matsuzawa, T. (2004). Color classification by chimpanzees (*Pan troglodytes*) in a matching-to-sample task. *Behav Brain Res, 148*(1-2), 157-165.

Matsuzawa, T. (1985). Colour naming and classification in a chimpanzee (*Pan troglodytes*). *J Hum Evol, 14*(3), 283-291.

Matsuzawa, T. (1990). Form perception and visual acuity in a chimpanzee. *Folia Primatol, 55*(1),

24-32.

Matsuzawa, T., Tomonaga, M., Tanaka, M. (2006). *Cognitive development in chimpanzees*. Tokyo: Springer.

Nahm, F. K. D., Perret, A., Amaral, D. G., Albright, T. D. (1997). How do monkeys look at faces? *J Cogn Neurosci, 9*(5), 611-623.

Parr, L. A., Dove, T., Hopkins, W. D. (1998). Why faces may be special: evidence of the inversion effect in chimpanzees. *J Cogn Neurosci, 10*(5), 615-622.

Parr, L. A., Hecht, E., Barks, S. K., Preuss, T. M., Votaw, J. R. (2009). Face processing in the chimpanzee brain. *Curr Biol, 19*(1), 50-53.

Shepherd, S. V., Steckenfinger, S. A., Hasson, U., Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr Biol, 20*(7), 649-656.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *J Vis, 7*(14).

Tomonaga, M. (2007). Visual search for orientation of faces by a chimpanzee (*Pan troglodytes*): face-specific upright superiority and the role of facial configural properties. *Primates, 48*(1), 1-12.

Tomonaga, M. (2010). Chimpanzee eyes have it? Social cognition on the basis of gaze and attention from the comparative-cognitive-developmental perspective. In E. Lornsdorf, S. Ross & T. Matsuzawa (Eds.), *The mind of the chimpanzee: Ecological and empirical perspectives*. Chicago: University of Chicago Press.

Tomonaga, M., Imura, T. (2008). Chimp in the shadow: Efficient detection of chimpanzee body by chimpanzees? *Primate Res, 24*(S), 14-15 (Japanese abstract only).

Tomonaga, M., Imura, T. (2009). Faces capture the visuospatial attention of chimpanzees(*Pan*

troglodytes): evidence from a cueing experiment. *Front Zool, 6*(1), 14.

Tomonaga, M., Matsuzawa, T. (1992). Perception of complex geometric-figures in chimpanzees (*Pan troglodytes*) and humans (Homo sapiens): Analysis of visual similarity on the basis of choice reaction time. *J Comp Psychol, 106*(1), 43-52.

Walther, D., Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395-1407.

Yarbus, A. L. (1967). *Eye Movements and Vision*. New York: Plenum Press.

Content Note

1. In the ANOVAs, in cases in which the assumption of homogeneity of variance was violated, the Greenhouse-Geisser correction was applied, and corrected *P* values were calculated.

Tables

Table 2-2-1. *Procedures Used for Image Manipulation*

condition	n	procedure
control	20	
monochrome	5	The color was removed from the original photographs.
line drawing	5	Only edges were extracted from the monochrome photograph (with a Photoshop function), and binary image processing techniques simplified the image (emphasizing the fat lines and eliminating the thin lines and small dots).
schematic drawing	5	The edges were roughly traced with simple black circles and lines.
blurred	5	The edges were blurred to the extent that the facial features were not recognizable (a Gaussian blur 20 pixels in diameter).
silhouette	5	The figure was colored in black, and binary image processing techniques transformed the background into black and white patches.
upside down	5	The original photographs were turned upside down.
scrambled	5	The original scenes were superimposed into a 6 × 5 matrix, and each block of the matrix was randomly rearranged. A matrix was defined so that a block includes the whole face (i.e. the face was intact).
headless	5	The head was eliminated so that the background was visible through the regions in which the head was previously located. To this end, the headless figure was cropped in the first image and superimposed on the second image that contains only background.

Figure captions

Figure 2-2-1. Scanpaths of a chimpanzee and a human, each superimposed on the naturalistic scene (a) and fine art painting (b; Paul Klee, 1923, "*Puppet Theater*"). Fixations and saccades are indicated by dots and lines, respectively. The stimuli were presented for 3 sec. each. Also shown are a raw saliency map and the scanpath predicted by the model. Bright areas indicate areas of high saliency. By design, the model made 9 fixations on the images in the order of decreasing saliency.

Figure 2-2-2. The locations of all fixations made by a chimpanzee, a human, and the model. While the model showed a relatively even distribution of fixations over the scene, the chimpanzee and the human showed a central bias in the distribution. Therefore, the chance level (random gaze pattern) was adjusted to control for this observed bias (see text).

Figure 2-2-3. (a) Proportion of fixations on each area of interest (AOI; see the diagram for an example) in each image by chimpanzees (n = 6) and humans (n = 16). (b) Proportion of images (n = 20) in which a fixation was observed in each AOI at each fixation order. The first 6 fixations are presented here. (c) The sum proportion of images at the first two fixations, showing that the results from (a) are evident no later than the first two fixations. The data are from the control condition. All data are shown as the difference from the chance level. T-tests compared between chimpanzees and humans, and between each species and the model (one-sample). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Error bars indicate s.e.m.

Figure 2-2-4. The saliency values at the first 6 fixations. The saliency value was standardized, and ranges from 0 (not salient) to 1 (highly salient). The data are taken from the control condition. n.s. not significant. Error bars indicate s.e.m.

Figure 2-2-5. (a) Examples of stimuli presented in each experimental condition. Note that the original stimuli were in color. (b) Proportion of fixations on the face in each image by chimpanzees (n = 6), humans (n = 16), and the model. (c) The mean saliency values at the first 6 fixations for

chimpanzees and humans. All data are shown as the difference from the chance level. T-tests compared between chimpanzees and humans, and between each species and the model (one-sample). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n.s. not significant (the P values for Figure 2-2-5c are 0.75, 0.20, 0.29, 0.78, 0.74, 0.19, 0.26, 0.86, 0.19, for each condition, from left to right). Error bars indicate s.e.m.

Figures

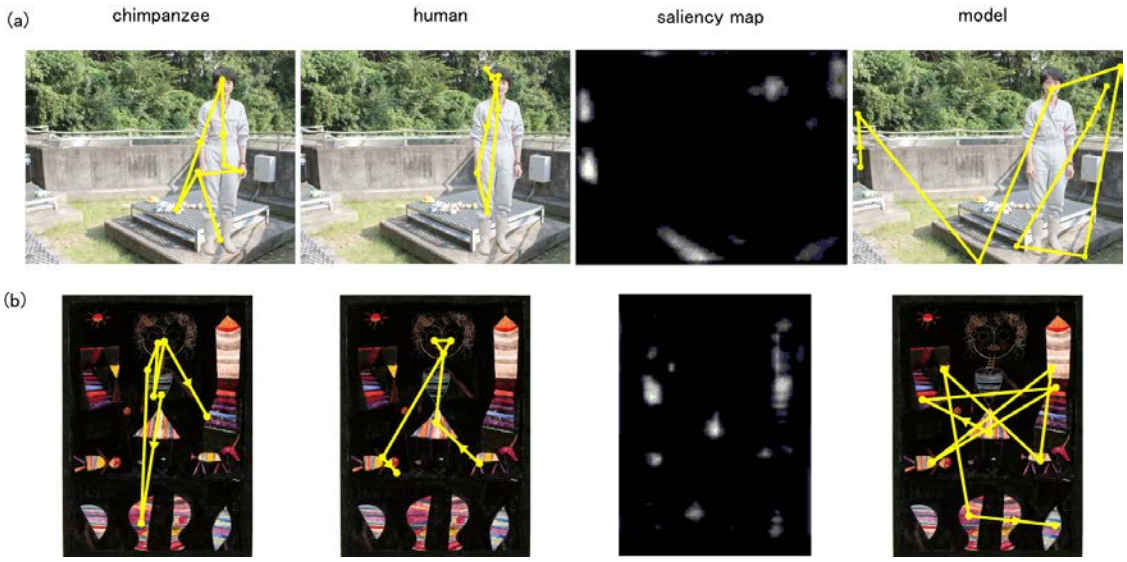


Figure 2-2-1

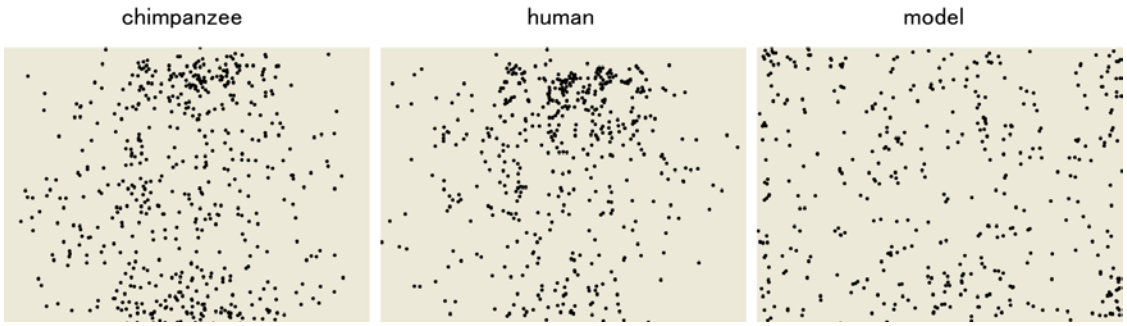


Figure 2-2-2

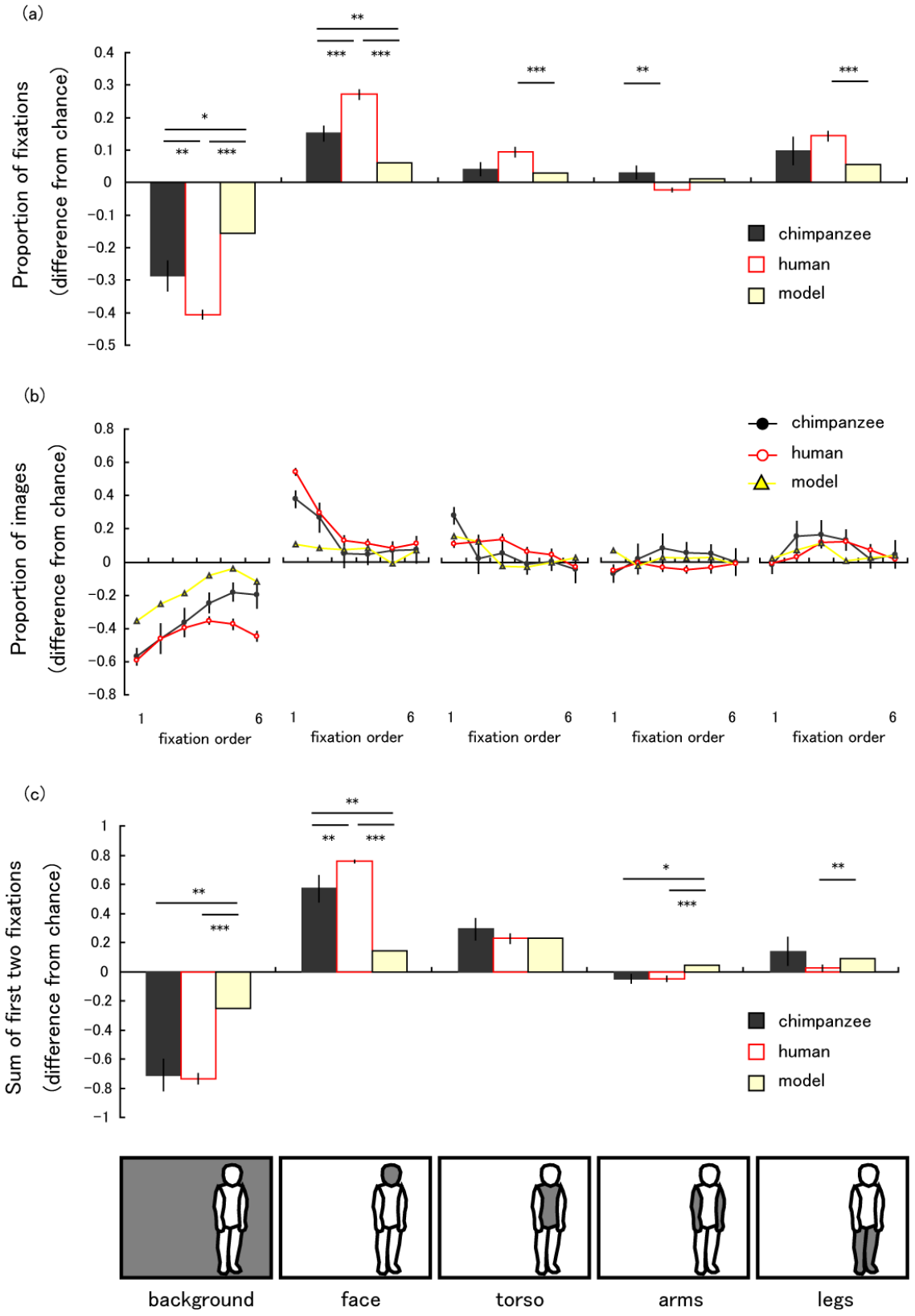

(a)
Proportion of fixations
(difference from chance)
chimpanzee
human
model
(b)
Proportion of images
(difference from chance)
fixation order
(c)
Sum of first two fixations
(difference from chance)
background
face
torso
arms
legs

Figure 2-2-3

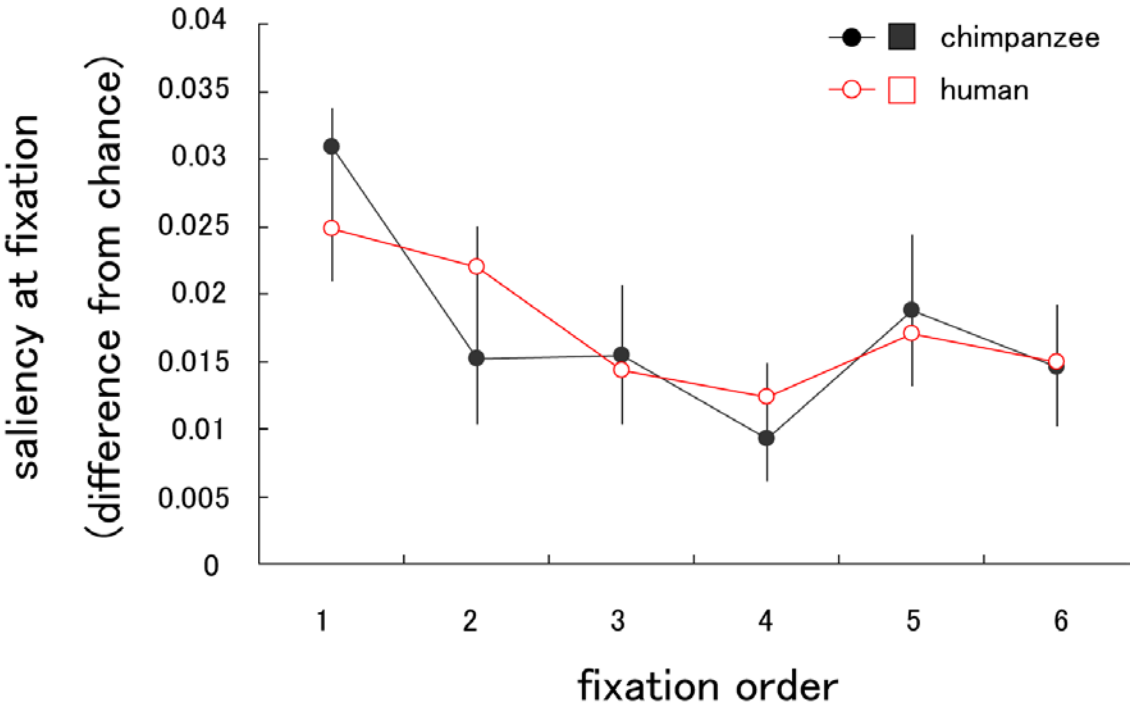

saliency at fixation
(difference from chance)
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
1
2
3
4
5
6
fixation order
chimpanzee
human

Figure 2-2-4

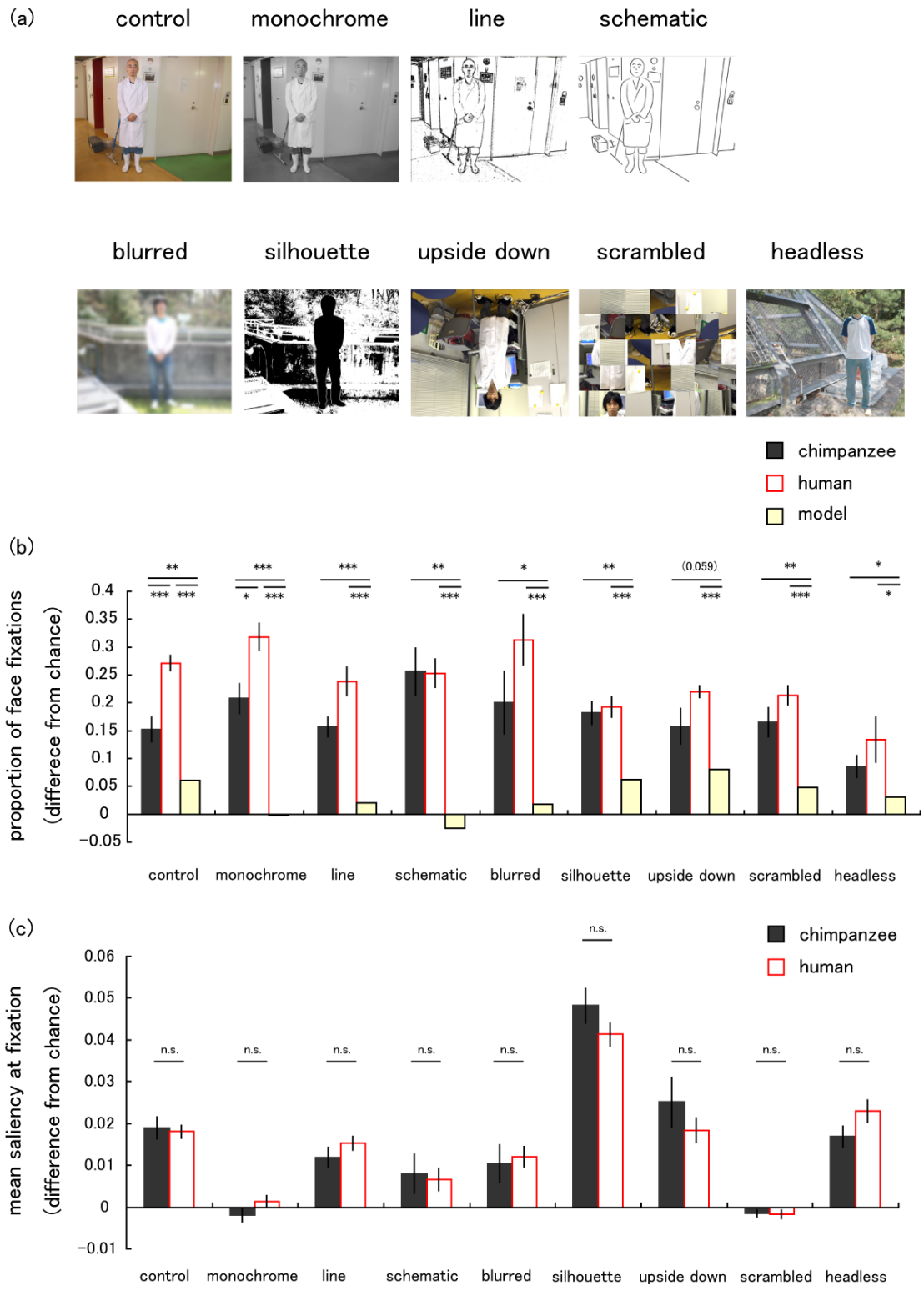

(a)
control
monochrome
line
schematic
blurred
silhouette
upside down
scrambled
headless
chimpanzee
human
model
(b)
proportion of face fixations
(differece from chance)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
−0.05
control
monochrome
line
schematic
blurred
silhouette
upside down
scrambled
headless
(c)
mean saliency at fixation
(difference from chance)
0.06
0.05
0.04
0.03
0.02
0.01
0
−0.01
n.s.
chimpanzee
human

Figure 2-2-5

Chapter 2-3.

Species difference in the timing of gaze movement between chimpanzees and humans

Introduction

The visual strategy common across human and nonhuman primates includes the alternation of fixation and saccade; by fixation, we keep certain parts of the visual field fixed on the fovea, where retinal acuity and colour sensitivity are optimized; by saccade, we bring new parts of the visual field onto the fovea using rapid eye movements. Although we have an impression that we see all parts of visual field with full clarity and resolution, in fact we obtain high-quality information only from the fovea. Therefore, a trade-off exists between the amount of time gaze is held still and the need to shift the gaze to attend to competing information that might be prioritized by the fovea for analysis. Hence, the timing of gaze movements reflects the way in which visual information is retrieved and analyzed.

Human adults normally make fixations at a frequency of up to three per second, and thus, the duration of each fixation lasts for up to 330 msec. However, the timing of gaze movements varies significantly around this mean within and across individuals (Henderson & Hollingworth, 1999). Studies have shown that the timing of gaze movements varies depending on ongoing cognitive processing activities (Findlay & Walker, 1999). For example, when reading text, humans increase the duration of their fixations as the text becomes more demanding semantically (Rayner, 1998). It is also known that individuals show different timings of gaze movement depending on their developmental stage and neurological condition. For example, human infants detach their gaze more slowly from the fixated object in response to peripheral stimuli, as they are in an earlier stage of development (Hood & Atkinson, 1993). Autistic children are also known to have difficulty in detaching their gaze from the fixated object compared with typically developing children (Landry & Bryson, 2004).

Several studies have been conducted in non-human primates. Studies have shown that during free viewing of a naturalistic dynamic scene, macaque monkeys shifted their gaze at an

earlier time than did humans and thereby scanned the scene more rapidly (Berg, Boehnke, Marino, Munoz, & Itti, 2009; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010). Kano & Tomonaga (2009) have found similar results in chimpanzees (*Pan troglodytes*), the species closest to humans. During free viewing of a naturalistic static scene, chimpanzees scanned the scene more rapidly than did humans. Therefore, humans may have a differential pattern for the timing of gaze movement compared with other primate lineages.

However, in contrast to the long history of eye-movement studies in macaques (e.g., Fuch, 1967), few studies have examined gaze dynamics in non-human great ape species. Therefore, studies are necessary to fill the phylogenetic gap between macaques and humans for the benefit of the comparative-evolutionary perspective on this topic. Given the established knowledge in the human and macaque literature showing that the timing of gaze movement reflects neurological and cognitive processes (e.g., Findlay & Walker, 1999), studies on primate groups closest to humans should provide insight into perceptual and cognitive evolution in human and non-human primates.

Chimpanzees are the closest species to humans genetically and are known to be similar to humans in a number of their visual abilities, such as visual acuity (Matsuzawa, 1990) and colour (Matsuno, Kawai, & Matsuzawa, 2004; Matsuzawa, 1985), form (Matsuzawa, 1990), and face perception (Parr, Dove, & Hopkins, 1998; Parr, Hecht, Barks, Preuss, & Votaw, 2009; Tomonaga, 2007; Tomonaga & Imura, 2009). Recent evidence also suggests that chimpanzees are comparable to humans in tests involving short-term memory and the functional visual field (Inoue & Matsuzawa, 2007).

Previous studies that compared the gaze movements of chimpanzees and humans using an eye-tracking system (Kano & Tomonaga, 2009, 2010, 2011) found that during free viewing of a naturalistic scene, the species were very similar in terms of the location of fixations over the scene (“where” to shift gaze). For example, they concentrated fixations on socially informative regions

such as faces of animals. A subsequent study (Kano and Tomonaga, 2011) found that chimpanzees and humans did not significantly differ in the extent to which low-level visual properties (e.g., colour, contrast, orientation) influenced their patterns of fixation distribution.

On the other hand, as mentioned above, these previous studies found a difference between chimpanzees and humans in the duration of each fixation ('when' to shift gaze); chimpanzees employed shorter fixations at each location and therefore scanned a scene more rapidly than did humans. However, although this is also the case for previous studies in macaque monkeys, the finding is somewhat confounded by the fact that the scene contained various kinds of local objects (face, body, tree, food, etc.) and that chimpanzees and humans showed different fixation durations for different kinds of local objects (e.g., longer fixation duration for faces in humans than in chimpanzees). Therefore, in this study, we prepared stimuli depicting various subject matters and asked whether the species difference reflects specific responses to the particular subject matter of the scenes/objects or more generalized responses to visual stimuli. In addition, it remains unclear whether the species difference in the fixation duration results from differential sensitivities to the peripherally-presented stimuli *per se* or from specific behavioural strategies used to deal with the competition between central and peripheral stimuli in their visual field (i.e., the trade-off between fixation and initiation of a saccade). We addressed this issue using a version of the gap–overlap paradigm, which was originally devised to examine visual development in human infants (Hood & Atkinson, 1993). Finally, we asked if there is a type of scene where the chimpanzees' gaze strategy works more beneficially to fixate local objects than does humans' gaze strategy. Specifically, we asked whether chimpanzees better track local objects at the point of fixation than do humans when presented with a dynamic scene where local objects appear/disappear successively in random positions at a chosen presentation rate.

General method

We conducted four experiments in this study. Chimpanzees and humans were tested using the same experimental procedure (with a few exceptions, see below) to enable a direct comparison. They were allowed to freely explore stimuli presented on a screen. Previous studies (Hattori, Kano, & Tomonaga, 2010; Hirata, Fuwa, Sugama, Kusunoki, & Fujita, 2010; Kano & Tomonaga, 2009, 2010) have established that eye-tracking recordings in chimpanzees and humans are both accurate and comparable (for the details, see below).

Subjects

Six chimpanzees (five females, one male; aged 8–31 years) and 18 humans (12 females, six males; aged 18–31 years) participated in Experiments 1, 2, and 4. Experiment 3 was conducted as a follow-up experiment to Experiment 2, and the same six chimpanzees and the same 10 humans participated. The chimpanzees were members of a social group consisting of 14 individuals living in an enriched environment with a 700-m^2 outdoor compound and an attached indoor residence (Matsuzawa, Tomonaga, & Tanaka, 2006). The outdoor compound was equipped with 15-m-high climbing frames, small streams, and various species of trees (Ochiai & Matsuzawa, 1997). Access to the outdoor compound was available to each individual every other day during the day. Daily meals included a wide variety of fresh fruits and vegetables fed throughout the day, supplemented with nutritionally-balanced biscuits (fed twice daily) and water available *ad libitum*. No deprivation of food or water was experienced by the chimpanzees during the study period. Care and use of the chimpanzees adhered to the 2002 version of the Guidelines for the Care and Use of Laboratory Primates of the Primate Research Institute, Kyoto University. The experimental protocol was approved by the Animal Welfare and Care Committee of the Institute and by the Animal Research Committee of Kyoto University. Informed consent was obtained from all human participants. The chimpanzees had extensive experience observing images displayed on a computer screen and

conducting perceptual and cognitive tasks using a touch-sensitive screen (Matsuzawa, Tomonaga, & Tanaka, 2006). However, neither chimpanzees nor humans had been explicitly trained for scanning scenes or shifting their gaze rapidly.

Apparatus

Both species used the same apparatus for the purpose of direct comparison between the species. Participants sat still and unrestrained in an experimental booth, and the eye-tracking apparatus and the experimenter were separated by transparent acrylic panels. A table-mounted eye tracker measured participants' eye movements using infrared corneal reflection techniques (60 Hz; Tobii X120, Tobii Technology AB, Stockholm, Sweden). This eye-tracker has wide-angle lenses (±40° in a semicircle above the camera) and thus obviated the necessity to restrain the subjects. The eye-tracker and the 17-inch LCD monitor (1280 × 1024) were mounted on a movable platform, and the distance between the platform and the participants was adjusted to the point at which the gaze was most accurately recorded (60 cm ± 10 s.e.m.). This flexible adjustment of the distance between the platform and the participants enabled us to record the gaze movements of chimpanzees with their heads unrestrained. The participant's gaze was recorded as a relative coordinate with respect to the monitor size (i.e., not as the gaze angle). One degree of gaze angle corresponds to approximately 1 cm on the screen at a typical 60-cm viewing distance.

As a result of the training conducted during the study performed by Kano and Tomonaga (2009), the chimpanzees were already skilled at sitting still in front of an eye-tracker and looking upon request at a fixation point that appeared on the screen. Five-point calibration was conducted for humans; for chimpanzees, the calibration points were reduced to two in order to decrease the time required for each calibration process. However, for chimpanzees, the calibration was repeated until the maximum accuracy was obtained. The accuracy was checked in both species by presenting five fixation points on the screen. Using these calibration procedures, six participants of each species were tested for accuracy, and the errors were found to be small and comparable between the species (average errors of 0.62 ± 0.06 and 0.52 ± 0.05 cm ± s.e.m. on the monitor for chimpanzees and humans, respectively). The drift (the calibration error due to changes occurring in the eye surface)

was checked occasionally by presenting the fixation points to the participants again.

Procedure

Procedural differences in testing chimpanzees and humans were minimized to allow for direct comparisons between the two species. In each trial, an image was presented after participants had focused on a fixation point that appeared in a central position on the screen. Participants were then allowed to freely view the images. Participants of both species rarely kept their gaze at the fixation point after the image presentation (i.e., spontaneous scanning was almost certain). Unlike some well-known reports on macaque eye tracking (Gothard, Erickson, & Amaral, 2004; Mendelson, Haith, & Goldmanrakic, 1982), we never observed fear responses to the images in the chimpanzees in either this or previous studies (Kano & Tomonaga, 2010). The entire session was conducted on a single day for humans, and was spread out over 12 days for the chimpanzees to decrease the time required for each experiment per day. On each day, the chimpanzees were presented with six stimuli. Daily experiments lasted 10–15 min for the chimpanzees and 20 min for humans. Following completion of the session, the human participants received book coupons as payment, whereas the chimpanzees received a small piece of apple after each trial. The reward was given to the chimpanzees in return for the initial fixation at the beginning of the trial and was provided independently of their viewing behaviour during the image presentation. Overall, these differences in procedure were designed to increase the motivation of both species to participate in the experiment from day to day and to sustain interest during the presentation of each image.

A fixation was scored if the gaze remained stationary within a radius of 50 pixels for at least 75 ms (more than five measurement samples). Otherwise, the recorded sample was defined as part of a saccade. We excluded samples recorded during the first 200 ms because it has been established that the spontaneous gaze-movement response to the stimuli occurs after that period in primates (as shown by Experiment 2).

Experiment 1

We first presented naturalistic scenes containing various kinds of subject matter (Figure 2-3-1b) including social (animal), ecologically meaningful (fruit), neutral (object), and meaningless (texture) stimuli to examine how chimpanzees and humans differ in their patterns of scanning the scenes and whether semantic categories might influence differences between the two species.

Method

Stimulus types consisted of animal (a scene containing several animals, either chimpanzees or humans), fruit (a scene containing several kinds of fruit), object (a scene containing several nonliving objects), and texture (a textured image of a natural object). Six stimuli were prepared for each stimulus type (24 stimuli in total). Stimuli measured 1000 × 750 pixels in size (26 × 20° at a typical 60-cm viewing distance) and were shown for 3 s. The presentation order of stimulus types was randomized for each participant. The area of interest (AOI) was defined with respect to the face regions in the animal scenes. To define AOI, polygons were drawn slightly (approx. 20 pixels) larger than the actual outlines to avoid errors in gaze estimation. After the whole session was completed, we repeated trials in which the gaze data had been lost for longer than 600 ms due to participants looking away from the monitor or blinking more than twice. We then replaced these trials with the new trials if those were completed satisfactorily; if not, we excluded these trials from the analysis. The data loss totalled 3.7% and 5.0 % of all trials for chimpanzees and humans, respectively (no bias for a particular stimulus type). Chimpanzees and humans viewed the monitor for equal durations during successful trials.

Results and Discussions

We found that chimpanzees scanned the scenes more rapidly; that is, the average fixation duration was shorter in chimpanzees than in humans (ANOVA, $F_{1, 22} = 16.68$, $p < 0.001$, $\eta^2 = 0.43$; Table 2-3-1, Figure 2-3-1c). Additionally, we established that this difference in average fixation time

was independent of the subject matter of the scene because there was no significant effect of stimulus type on average fixation duration ($F_{3, 66}$ = 1.13, p = 0.34, η^2 = 0.04) or on the species difference (species × stimulus type, $F_{3, 66}$ = 0.87, p = 0.46, η^2 = 0.03). This does not imply that chimpanzees and humans neglected the subject matter of the scenes, as the fixations of both species were concentrated on socially informative regions, such as faces in the animal category (Figure 2-3-2). Both chimpanzees and humans showed a high proportion of face fixations (average 4.7 ± 0.47 and 4.5 ± 0.20 fixations on the face (± s.e.m.); 49 ± 5.3 and 63 ± 3.0 % (± s.e.m.) in all scene fixations, respectively). Therefore, it was concluded that the observed species difference in the duration of fixations reflected the general tendency employed by each species to scan the visual scenes.

We also confirmed that chimpanzees' engaging in shorter fixations indicated that they scanned the scene more widely than did humans during the presentation period (Figures 2-3-1a and 2-3-2). This was established by dividing the scene area into 8 × 6 grids and counting the number of grids covered by the fixations (Table 2-3-1). Because the number of repeated fixations on each grid was the same for chimpanzees and humans (Table 2-3-1), the shorter fixations by chimpanzees (thus, larger numbers of fixations) allowed fixation on a greater number of cells by the chimpanzees compared with the humans. The benefit to chimpanzees of scanning scenes this way was further examined in Experiment 4.

Figure 2-3-1

Figure 2-3-2

Experiment 2

In this experiment, we hypothesized that the species difference in fixation duration was due to a differential strategy for coping with the competition between the two activities of fixation and initiation of a saccade. That is, chimpanzees engage in shorter fixations because they are bias toward the latter rather than the former in their strategy and thus initiate a saccade sooner than do humans (hypothesis 1). A possible alternative explanation is that, compared with humans, chimpanzees might perceive peripheral visual stimuli as more salient and thus initiate a saccade in a more sensitive way (hypothesis 2), although this explanation seems unlikely given the similarity in the perceptual capacities of the two species (see above). To test these hypotheses, we used a version of the gap–overlap paradigm, which was originally devised to examine visual development in human infants (Hood & Atkinson, 1993) (Figure 2-3-3a). Two conditions, gap and overlap, were employed. Under these conditions, a central (fixated) and peripheral target stimulus appeared in that order (Figure 2-3-3a). Under the gap condition, the central fixation stimulus disappeared shortly before target presentation, but under the overlap condition, the central fixation stimulus remained. The time between target presentation and initiation of a saccade to the target was measured (i.e., saccade latency). It is well known that in humans, saccade latency to peripheral stimuli tends to be longer under the overlap compared with the gap condition due to the presence of competition between the two activities involving fixation and initiation of a saccade under the overlap condition (known as the “gap effect”) (Findlay & Walker, 1999). Findlay & Walker’s model, which is based on extensive evidence from macaque and human studies, assumes that resolving such competition is a time-consuming process and that various perceptual and cognitive events occurring centrally or peripherally (such as processing of central stimuli and perceiving the sudden appearance of peripheral stimuli) influence this process at various levels from automatic, automated (or habitual), to voluntary. During competition between fixation and initiation of a saccade, a saccade is generated

when the latter overcomes the former. Therefore, in this study, if the species differed in their way of dealing with competition between central and peripheral stimuli (i.e., activities involving fixation and those involving initiation of a saccade), it is expected that they will differ in the degree of gap effect (no matter what level of process is involved; hypothesis 1). If, instead, the species difference results from their differential sensitivities for the peripheral target *per se*, it is expected that chimpanzees and humans will differ in the saccade latency under the gap condition (hypothesis 2).

Figure 2-3-3

In Experiment 2, two stimulus types, face and object, were prepared to separate the effect of stimulus type on the saccade latency from that of condition (gap and overlap). Face and object were treated as different stimulus types because these two stimulus types were attended to differently in both species, as shown in Experiment 1 and previous studies (Kano & Tomonaga, 2009; Tomonaga & Imura, 2009); faces were predominantly fixated when presented simultaneously with objects within a scene. The same stimulus type (either face or object) was presented at central and peripheral locations within a trial. We expected that species difference would appear for both types of stimuli, following the results obtained in Experiment 1.

Method

A central fixation stimulus and a target stimulus (approx. 180 × 180 pixels) were presented in that order (approx. 340 pixels apart from each other; 9° at a typical 60-cm viewing distance), and the target was presented randomly to the left or right 560 ms after the onset of the trial (Figure 2-3-3a). The time between target presentation and initiation of a saccade to the target was measured (i.e., saccade latency). Under the gap condition, the central fixation stimulus disappeared 260 ms before target presentation, but under the overlap condition, the central fixation stimulus remained.

Each trial lasted for 1.5 s in total. The stimulus type was either face or object within a trial. More than 50 exemplars each were prepared for face (equal number of exemplars for chimpanzee or human faces) and object, and different exemplars were used for central and peripheral locations within a trial. Each exemplar was randomly selected from the entire exemplar pool. Six trials were prepared for each condition and stimulus type (24 trials in total). The presentation order of conditions and stimulus types was randomized for each participant. After the completion of the whole session, we repeated trials in which participants shifted their gaze before the presentation of the peripheral target or failed to shift their gaze to the target. We then replaced them with new trials if those were satisfactory; if not, we excluded these trials from the analysis. The data loss totalled 2.7% and 1.1% of all trials for chimpanzees and humans, respectively (no bias for a particular stimulus type or condition).

Results and Discussion

As previously reported, the saccade latency was found to be longer under the overlap than under the gap condition in humans (Figure 2-3-3b; 106 ms, $F_{1, 17} = 43.81$, $p < 0.001$, $\eta^2 = 0.72$). In chimpanzees, however, this effect was much smaller (17 ms, $F_{1, 5} = 9.07$, $p = 0.030$, $\eta^2 = 0.64$), resulting in a significantly greater gap effect in humans compared with chimpanzees (species × condition, $F_{1, 22} = 9.52$, $p = 0.005$, $\eta^2 = 0.30$). Moreover, a significant species difference was found under the overlap condition ($F_{1, 22} = 8.18$, $p = 0.009$, $\eta^2 = 0.27$), but not under the gap condition ($F_{1, 22} = 0.95$, $p = 0.33$, $\eta^2 = 0.04$). In these results, no significant effect of stimulus type was found ($p > 0.05$).

Therefore, the species difference appeared in the difference between the gap and overlap conditions (i.e., gap effect) rather than in the gap condition, supporting hypothesis 1. That is, chimpanzees and humans differ in their behavioural strategy for dealing with the competition between central and peripheral visual stimuli in their visual field rather than in their sensitivity to the

peripheral stimuli *per se*. In Experiment 2, we did not observe any significant effect of stimulus type. This may be because the central and peripheral stimuli of the same stimulus type competed equally with each other and, thus, did not strongly influence the saccade latency. An alternative explanation is that face and object were attended similarly by both species, although this is unlikely given the results obtained in Experiment 1. Experiment 3 was conducted to eliminate this possibility more thoroughly.

Experiment 3

In Experiment 3, a face was presented centrally along with an object at the periphery (face–object pair) or *vice versa* (object–face pair). As faces are known to attract the attention of both species more strongly than objects (see above), the competition is expected to be biased centrally in the former and peripherally in the latter pair. We thus expected that saccade latency would increase in the former and decrease in the latter pair. A critical question with regard to species difference is whether the species difference in the gap effect (i.e., an interaction between species and condition) would remain even when the competition between central and peripheral stimuli was experimentally biased in such a way.

Method

In Experiment 3, a face was presented centrally along with an object image at the periphery (face–object) or *vice versa* (object–face). Six trials were prepared for each condition (gap and overlap) and stimulus pair (face-object and object-face; 24 trials in total). The other procedures were the same as in Experiment 2. The data loss totalled 3.4% and 1.3 % of all trials for chimpanzees and humans, respectively (no bias for a particular stimulus pair or condition).

Results and Discussion

As in Experiment 2, a significant interaction between species and condition (gap and overlap) was found (Figure 2-3-4; $F_{1, 13} = 4.66$, $p = 0.050$, $\eta^2 = 0.264$). In addition, a significant main effect of stimulus pair was found ($F_{1, 13} = 14.25$, $p = 0.002$, $\eta^2 = 0.52$). However, no interaction was found between stimulus pair and species, condition, or both ($p > 0.05$).

Therefore, we confirmed that the main effect of stimulus pair appears but that this effect occurs relatively independently of the species difference in the gap effect, not as an interaction with species or condition. We also confirmed that a face dominates in the competition between central and peripheral vision in both species. Thus, the absence of stimulus-type effects in Experiment 2 was

indeed due to the fact that central and peripheral locations competed equally with each other. Nonetheless, the robustness of the species effect against manipulation of stimulus type is a little surprising given that the effect of a particular stimulus type (e.g., face) on the saccade latency was also robust over the others; further study is necessary to determine whether the manipulation of stimulus type offsets the effect of species and condition. One possible explanation of this robust effect of species is that the species difference involves competition occurring at a relatively lower level of processing (e.g., automatic or well-automated perceptual/cognitive process in Findlay and Walker's model).

Figure 2-3-4

Experiment 4

Experiments 2 and 3 showed that chimpanzees shifted their gaze more rapidly than did humans to the peripheral visual stimuli when competition between central (fixated) and peripheral stimuli was present in their visual fields. Therefore, the visual strategy of chimpanzees may have benefits over that of humans in terms of their speed of scanning. This issue was also pointed out in Experiment 1; chimpanzees fixated each location for less time and thereby scanned a wider area of the visual scene more rapidly than did humans. Experiment 4 set out to show this issue more clearly under a controlled condition by modifying the task used in Experiment 2. We increased the number of local targets in the image (Figure 2-3-5a), and these targets appeared/disappeared successively at random locations on the screen.

Method

At intervals of 260 ms, small images (approx. 140 × 140 pixels; 3.7 × 3.7° at a typical 60-cm viewing distance) were presented at random locations on the screen (within an area of 1000 × 750 pixels; 26 × 20° at a typical 60-cm viewing distance; no overlap in location for items) for 700 ms. The presentation lasted for 3 s and consisted of a total of 12 objects. Two stimulus types, face and object, were prepared. The same stimulus type (either face or object) was presented within a trial. The same exemplars as those in Experiments 2 and 3 were used, and a different exemplar was presented at each location within a trial. Each exemplar was randomly selected from the entire exemplar pool. Twelve trials were prepared for each stimulus type (24 trials in total). The presentation order of stimulus types was randomized for each participant. We repeated trials in which the gaze data were lost for longer than 600 ms due to participants' looking away from the monitor or blinking more than twice after the completion of the whole session. We replaced them with new trials if those were completed satisfactorily; if not, we excluded these trials from the analysis. The data loss totalled 2.7% and 0.0% of all trials for chimpanzees and humans, respectively

(no bias for a particular stimulus type). In the successful trials, chimpanzees and humans viewed the monitor for equal durations.

Results and Discussion

Under this condition, chimpanzees shifted their gaze sooner (i.e., fixated more briefly) and thereby fixated on more images than did humans (Table 2-3-1, Figure 2-3-5b; $F_{1, 22} = 14.95$, $p = 0.001$, $\eta^2 = 0.40$). It was additionally confirmed that chimpanzees viewed the fixated images for longer periods compared with humans ($F_{1, 22} = 12.55$, $p = 0.002$, $\eta^2 = 0.36$). This suggests that, compared with chimpanzees, humans actually spent more time fixating on a grey background. In these results, no significant effect of stimulus type was found ($p > 0.05$), which is consistent with Experiments 1 and 2.

These results replicated (and combined) the results obtained in Experiments 1 and 2. A somewhat surprising result came from the fact that chimpanzees not only fixated a wider variety of images but also fixated those images for longer periods than did humans. That is, chimpanzees held those images in the fovea for longer periods than did humans.

It should be noted that in Experiments 1–4, the conditions were prepared to minimize the effects of instruction or training on the participants' control of gaze movements. Therefore, participants were not provided with instructions, and in Experiments 2–4, the location at which a target might appear was not predictable. Furthermore, the gaze movements of participants did not differ significantly throughout the sessions in Experiments 1–4. This issue was examined by dividing the whole session into four blocks; integrating the blocks into the analysis for each dependent variable in Experiments 1–4 showed no effect of block (ps > 0.05).

Figure 2-3-5

General Discussion

In this study, we directly compared the gaze movements of humans and the closest species, chimpanzees, using an eye-tracking system. During free viewing of a naturalistic scene, chimpanzees made more fixations per second (up to four) than did humans (up to three). This species difference was independent of the semantic variability of the presented scenes (Experiment 1). The gap–overlap paradigm revealed that, rather than resulting from the differential sensitivities to the peripherally-presented stimuli *per se*, the species difference reflected the particular strategy used by each species to solve the rivalry between central (fixated) and peripheral stimuli in their visual fields (Experiment 2). Again, this species difference was independent of the variability of stimulus types (Experiments 2 and 3). Finally, when presented with a movie where small images appeared/disappeared at random positions successively, chimpanzees retained those images at the point of fixation for a longer time than did humans, outperforming humans in their speed of scanning (Experiment 4). Our results demonstrate that chimpanzees and humans differ quantitatively in their visual strategies involving the timing of gaze movement (Figs. 7, 8).

In this study, we were interested in the spontaneous ("natural" or "habitual") gaze movement of chimpanzees and humans rather than their ability to control gaze movements. Thus, first, chimpanzees were rewarded independently of their viewing behaviour, i.e., even when they did not view the screen (although chimpanzees spontaneously viewed the screen in most of the trials; see Method). Second, in both species, heads were not restrained so that participants were not forced to view the screen. Third, both species received no instruction or training for scene scanning. Additionally, neither had been trained for gaze movements (except fixation training mentioned above) before this study. Fourth, the gaze movements of participants did not differ significantly throughout the sessions, showing no indication of effects of uninstructed training. Finally, the chimpanzees were not in a heightened state of vigilance because chimpanzees showed no fear

response to the presented images, even in the first trial, unlike in other well-known reports on monkeys (Gothard et al., 2004; Mendelson et al., 1982). Furthermore, chimpanzees were not more likely to be distracted by the environment outside of the monitor than were humans because both species shifted their attention to events occurring on the monitor (as seen in Experiment 4 particularly). Although these results suggest that chimpanzees and humans viewed the stimuli comparably and spontaneously, we cannot thoroughly exclude the possibility that chimpanzees and humans interpreted the current task (free-viewing) in a different way. To eliminate this possibility more thoroughly, further studies are necessary utilizing a more naturalistic setting (e.g., head-mount eye-tracking; (Land, Mennie, & Rusted, 1999)). It would also be worthwhile to train humans to move their gaze more frequently (like chimpanzees) and to observe the effect of training on their overall perceptual and cognitive abilities.

Do the differential gaze strategies of chimpanzees and humans reflect their differential evolutionary strategies? In answering this question, we consider, first, the rearing histories of participants from both species. Our chimpanzee participants have lived in a safe captive environment from a young age, with a secure supply of food. Thus, it is unlikely that differences in the severity of the rearing environment have shaped the differential patterns of gaze movement between the two species. Second, with respect to cultural effects on the pattern of gaze movement, it has been reported that those from Eastern cultures scanned the scenes more broadly and thus somewhat more rapidly than did those from Western cultures (Chua, Boland, & Nisbett, 2005). As our human participants were from an Eastern culture (Japan), it is unlikely that their cultural experience had a critical effect on the difference between the species. Third, as for the effect of task experience on the pattern of gaze movements, it has been reported that video action game players have an improved visual spatial resolution compared with non-players in a task involving discriminating symbols in their peripheral vision (Green & Bavelier, 2007). As our chimpanzee participants were highly

experienced using computer screens, such an effect of task experience may have emerged more strongly than it would have with other groups of chimpanzees or, possibly, with the human participants (nonetheless, they had never been explicitly trained for scanning the scenes rapidly). However, as our data suggest that chimpanzees and humans differ in their strategies for dealing with the competition between central and peripheral stimuli, but not in their sensitivities to the peripheral stimuli *per se*, this evidence does not seem critical to our conclusion regarding species difference. However, to more thoroughly control for the effect of task experience, it is necessary to compare chimpanzee groups that differ in their experiences of computer screen use.

At least three lines of research are worth pursuing in future studies, and these relate to the development, mechanism, and phylogeny of gaze-movement patterns. First, as mentioned in the Introduction, human infants detach their gaze more slowly from a fixated object in response to peripheral stimuli, as they are in an earlier stage of development; i.e., there is larger gap effect (Hood & Atkinson, 1993). In this study, chimpanzees showed an opposite tendency from human infants in this regard. Thus, it might be the case that humans acquired a neotenous pattern through development and evolution. Future studies might valuably compare the development of gaze-movement patterns in both chimpanzees and humans to observe when those two species start to show differences. Second, autistic children are known to have a difficulty in detaching their gaze from fixated objects (Landry & Bryson, 2004), a tendency opposite that of the chimpanzees in this study. It would, therefore, be worthwhile to conduct further studies to clarify the mechanism underlying such differences in the timing of gaze shift. Finally, it has been reported that macaque monkeys scan dynamic scenes more rapidly and broadly than do humans (Berg, Boehnke, Marino, Munoz, & Itti, 2009; Shepherd et al., 2010), similar to chimpanzees. Therefore, we suspect that the strategy involving the rapid scanning of a scene is a relatively common among non-human primates and that humans have acquired a gaze-movement strategy that is divergent from that of their

ancestors; further studies are necessary.

In Experiment 4, chimpanzees fixated on each local image better than humans did and thus tracked those images in the fovea for a longer period. There seemed to be at least two functional interpretations for the superiority of chimpanzees' gaze strategy over that of humans in terms of the speed of scanning. First, quick scanning of the scene may result from chimpanzees' exploring their species-specific visual environment. For example, chimpanzees habitually live in dense forest where the appearance of conspecific or dangerous animals is less predictable. In addition, chimpanzees live in social groups where the hierarchy among individuals and the competition over foods is occasionally severe (e.g., Pusey et al., 2008). Therefore, the frequent scanning of wider area in a scene may be more beneficial for chimpanzees than for humans. It would be valuable for future studies to test primate species from various ecological and social backgrounds for their patterns of gaze movement.

Second, rather than adaptation to specific environments, the species difference in gaze movements may reflect alternative strategies in which each species copes with its specific limitations with regard to retrieving and processing visual information. First, we ask if chimpanzees and humans retrieve visual information differently while they are fixating their eyes (i.e., without eye movement) and thereby employ different gaze-movement strategies. There are two possibilities. First, it is possible that humans have higher resolution in their peripheral field than chimpanzees have, and humans thereby prefer to retrieve visual information from the peripheral field and keep their gaze more static. However, this seems unlikely because in Experiments 2 and 3, we observed similar timings of gaze shifts between the species in the gap condition (i.e., the condition that measures the viewer's sensitivity to the peripheral stimuli *per se*). In addition, when given the opportunity to freely explore an image, chimpanzees and humans showed a similar pattern of fixation distribution across the scene (Experiment 1; also see (Kano & Tomonaga, 2009)). Furthermore, in Inoue &

Matsuzawa (2010), chimpanzees and humans were presented with multiple numeric symbols tachistoscopically (for less time than the duration of a typical fixation) in both central and peripheral visual fields and were subsequently required to recognize the numerical order. Chimpanzees and humans showed comparable accuracy on the task (some chimpanzees were even better than human participants). Second, it is possible that humans retrieve visual information more slowly from their fovea than do chimpanzees and therefore need to hold their gaze at the fixating spot for longer. This seems unlikely because within the typical period of human fixation (250–300 ms), less than 100 ms is spent retrieving the visual information necessary for object/word recognition (Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981; van Diepen, Wampers, & d'Ydewalle, 1998), and the remainder of the fixation time is therefore spent processing the retrieved information.

Thus, it is possible that chimpanzees and humans process the retrieved information differently and thereby employ different gaze-movement strategies. In humans, when engaging in text-reading (Rayner, 1998) and a visual-search task (Gould, 1973; Hooge & Erkelens, 1999), fixations tend to be longer as the processing load for foveal information becomes more demanding. In addition, when viewing a scene, it has been established that fixations tend to become longer as the fixated object contains more semantic information (De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, Weeks Jr. &, Hollingworth, 1999). Therefore, we speculate that a prolonged fixation duration in humans reflects the time allocated to higher-level processing (e.g., information integration, categorical processing, language processing). We suggest that due to certain limitations in processing speed (i.e., no greater ability to process visual information in humans than in chimpanzees), humans strategically retrieve a limited amount of visual information by keeping their gaze relatively static and thereby conserve time for information processing. However, although our data are consistent with this hypothesis, in this study, we did not directly examine the way in which chimpanzees and humans process (e.g., recognize and integrate) the retrieved information. This is a

limitation of the present study that should be resolved in future studies.

Our hypothesis, which involves the cost-benefit relationship between information retrieval and high-level information processing, is similar to (but not the same as) those proposed by Humphrey (2002) and Matsuzawa (2009) (short-term memory vs. symbolic representation, in their cases). An important message from these hypotheses is that increased brain size in humans may not necessarily have led to a general enhancement of perceptual/cognitive abilities, and certain types of processing interference may appear more strongly in humans than in their closest species, chimpanzees. Further studies are necessary to examine this issue.

In conclusion, this study directly compared the gaze movements of chimpanzees and humans using an eye-tracking system and found that the species differ in the timing of their gaze shift when scanning a visual scene. As gaze movement is known to be a sensitive indicator of cognitive processes and neurological conditions, it would be worthwhile to examine this finding more thoroughly in the future.

Figure 2-3-6

Figure 2-3-7

References

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *J Vision, 9*(5), 1-15.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *P Natl Acad Sci USA, 102*(35), 12629-12633.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychol Res, 52*(4), 317-329.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behav Brain Sci, 22*(04), 661-674.

Fuchs, A. F. (1967). Saccadic and smooth pursuit eye movements in the monkey. *J Physiol, 191*(3), 609-631.

Gothard, K. M., Erickson, C. A., & Amaral, D. G. (2004). How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Anim Cogn, 7*(1), 25-36.

Gould, J. D. (1973). Eye movements during visual search and memory search. *J Exp Psychol, 98*(1), 184-195.

Green, C. S., Bavelier, D. (2007). Video-action-game experience alters spatial resolution of vision. *Psychol Sci, 18*(1), 88-94.

Hattori, Y., Kano, F., & Tomonaga, M. (2010). Differential sensitivity to conspecific and allospecific cues in chimpanzees and humans: a comparative eye-tracking study. *Biology Lett*.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annu Rev Psychol, 50*, 243-271.

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *J Exp Psychol Hum Percept Perform, 25*(1), 210-228.

Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., & Fujita, S. (2010). Facial perception of conspecifics: chimpanzees (*Pan troglodytes*) preferentially attend to proper orientation and open eyes. *Anim Cogn, 13*(5), 679-688.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behav Dev, 16*(4), 405-422.

Hooge, I. T. C., & Erkelens, C. J. (1999). Peripheral vision and oculomotor control during visual search. *Vision Res, 39*(8), 1567-1575.

Humphrey, N. (2002). The deformed transformed. In *The Mind Made Flesh: Essays from the Frontiers of Psychology and Evolution* (pp. 165-199). New York: Oxford University Press.

Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Curr Biol, 17*(23), 1004-1005.

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: a comparative eye-tracking study. *Proc Roy Soc B, 276*(1664), 1949-1955.

Kano, F., & Tomonaga, M. (2010). Face scanning in chimpanzees and humans: continuity and discontinuity. *Anim Behav, 79*, 227-235.

Kano, F., & Tomonaga, M. (2011). Perceptual mechanism underlying gaze guidance in chimpanzees and humans. *Anim Cogn, 14*(3), 377-386.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311-1328.

Landry, R., & Bryson, S. E. (2004). Impaired disengagement of attention in young children with autism. *J Child Psychol Psychiat, 45*(6), 1115-1122.

Matsuno, T., Kawai, N., & Matsuzawa, T. (2004). Color classification by chimpanzees (*Pan troglodytes*) in a matching-to-sample task. *Behav Brain Res, 148*(1-2), 157-165.

Matsuzawa, T. (1985). Colour naming and classification in a chimpanzee (*Pan troglodytes*). *J Hum*

Evol, 14(3), 283-291.

Matsuzawa, T. (1990). Form perception and visual acuity in a chimpanzee. *Folia Primatol, 55*(1), 24-32.

Matsuzawa, T. (2009). Symbolic representation of number in chimpanzees. *Curr Opin Neurobiol, 19*(1), 92-98.

Matsuzawa, T., Tomonaga, M., & Tanaka, M. (2006). *Cognitive development in chimpanzees*. Tokyo: Springer.

Mendelson, M. J., Haith, M. M., & Goldmanrakic, P. S. (1982). Face scanning and responsiveness to social cues in infant rhesus monkeys. *Dev Psychol, 18*(2), 222-228.

Parr, L. A., Dove, T., & Hopkins, W. D. (1998). Why faces may be special: evidence of the inversion effect in chimpanzees. *J Cogn Neurosci, 10*(5), 615-622.

Parr, L. A., Hecht, E., Barks, S. K., Preuss, T. M., & Votaw, J. R. (2009). Face processing in the chimpanzee brain. *Curr Biol, 19*(1), 50-53.

Pusey, A., Murray, C., Wallauer, W., Wilson, M., Wroblewski, E., & Goodall, J. (2008). Severe aggression among female *Pan troglodytes schweinfurthii* at Gombe National Park, Tanzania. *Int J Primatol, 29*(4), 949-973.

Rayner, K. (1998). Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychol Bullet, 124*, 372-422.

Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *J Exp Psychol Hum Percept Perform, 7*(1), 167-179.

Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human–monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr Biol, 20*(7), 649-656.

Tomonaga, M. (2007). Visual search for orientation of faces by a chimpanzee (*Pan troglodytes*): face-specific upright superiority and the role of facial configural properties. *Primates, 48*(1), 1-12.

Tomonaga, M., & Imura, T. (2009). Faces capture the visuospatial attention of chimpanzees (*Pan troglodytes*): evidence from a cueing experiment. *Front Zool, 6*(1), 14.

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: moving masks and moving windows. In G. D. M. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337-355). Oxford: Elsevier.

Tables

Table. 2-3-1 Species differences in eye-movement characteristics between chimpanzees and humans (standard deviations by trial and individual variance).

	chimpanzees		humans		difference		*t*	*P*
Exp. 1								
average fixation duration (ms)	234	(20.0; 16.9)	323	(32.3; 67.9)	89	**	3.07	0.006
total number of fixations	9.8	(1.08; 0.43)	7.8	(0.65; 0.95)	2	***	4.88	< 0.001
number of grids fixated	7.4	(1.15; 0.50)	6.1	(0.68; 0.84)	1.3	**	3.68	0.001
average numbers of fixations on each grid	1.31	(0.14; 0.06)	1.28	(0.08; 0.09)	0.03		0.74	0.46
saccade size (pixel)	252	(37.8; 12.5)	241	(26.5; 30.7)	11		0.86	0.39
Exp. 2								
response time under gap condition (ms)	216	(10.9; 22.7)	229	(12.2; 29.0)	13		1.00	0.32
response time under overlap condition (ms)	232	(20.5; 28.3)	334	(45.4; 82.2)	102	**	2.92	0.008
gap effect (ms)	16	(20.8; 11.8)	105	(47.6; 67.6)	89	**	3.14	0.005
Exp. 3								
response time under gap condition (ms)	187	(19.9; 16.9)	202	(24.0; 39.7)	15		0.91	0.38
response time under overlap condition (ms)	186	(19.2; 20.4)	296	(63.6; 100.7)	110	*	0.22	0.02
gap effect (ms)	-1	(25.3; 18.0)	94	(55.7; 91.9)	95	*	0.17	0.027
Exp. 4								
number of images fixated	9.3	(0.38; 0.76)	6.1	(0.61; 1.97)	3.2	***	3.86	0.001
total viewing time for images (ms)	2051	(173; 144)	1498	(128; 368)	553	**	3.54	0.002
average fixation duration (ms)	213	(11.6; 22.0)	294	(22.5; 72.3)	81	*	2.67	0.014
total number of fixations	11.4	(0.39; 1.01)	8.8	(0.47; 1.54)	2.6	**	3.78	0.001
average number of fixations on each image	1.08	(0.02; 0.03)	1.10	(0.04; 0.08)	0.02		0.94	0.35
saccade size (pixel)	405	(32.1; 15.8)	337	(29.8; 42.5)	68	**	3.71	0.001

Note. df = 22 for Exp. 1, 2, 4; df = 13 for Exp. 3. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Figure captions

Figure 2-3-1. Species differences in fixation duration. a. Scan patterns in a chimpanzee and a human during free viewing, superimposed on the presented scene (toned down for clarity). The circles represent fixations (scaled in size to their durations, which are shown in milliseconds), and the lines represent saccades. The picture was presented for 3 s. b. Examples of still images presented in Experiment 1. c. Average fixation duration (ms) for chimpanzees (n = 6) and humans (n = 18). Error bars denote the standard error of the mean (s.e.m.).

Figure 2-3-2. Spatial distribution of fixations by chimpanzees and humans for an animal scene. Fixations by six members of each species are superimposed on the presented scene (toned down for clarity).

Figure 2-3-3. Species differences in saccade latency. a. Gap–overlap paradigm in Experiment 2. b. Saccade latency (ms) under the gap and overlap conditions. Error bars denote s.e.m.

Figure 2-3-4. The effect of stimulus pair on saccade latency (ms) in the gap–overlap paradigm (Experiment 3). Error bars denote s.e.m.

Figure 2-3-5. Species differences in object-tracking performance. a. Sequence from a video presented in Experiment 4 (also see SI for the examples). b. Number of fixated images (the number of images on which participants fixated at least once) and total viewing time (ms) for the images (the sum of the durations of fixations on any image). Error bars denote s.e.m.

Figure 2-3-6. Species differences in the frequency distribution of fixations (Experiments 1, 4; a, c, respectively) and saccade latency (Experiment 2; b). The data were pooled for all participants and stimuli. As previously reported for humans, chimpanzees and humans showed a skewed distribution in fixation durations, with the mode between 200 and 300 ms. The distribution of fixation durations was skewed more leftward in chimpanzees than in humans, and fixations over 300 ms were more frequent in humans than in chimpanzees. As shown in Figure b, the distribution of saccade latency

was similar for chimpanzees and humans under the gap condition. Under the overlap condition, however, the distribution was skewed more leftward in chimpanzees than in humans.

Figure 2-3-7. Individual and species differences in the timing of gaze movements (Experiments 1, 2, and 4). From a to f, Pearson's *r* was 0.64 ($P = 0.001$), 0.18 ($P = 0.39$), 0.49 ($P = 0.014$), 0.66 ($P < 0.001$), 0.81 ($P < 0.001$), and 0.83 ($P < 0.001$), respectively.

Figures

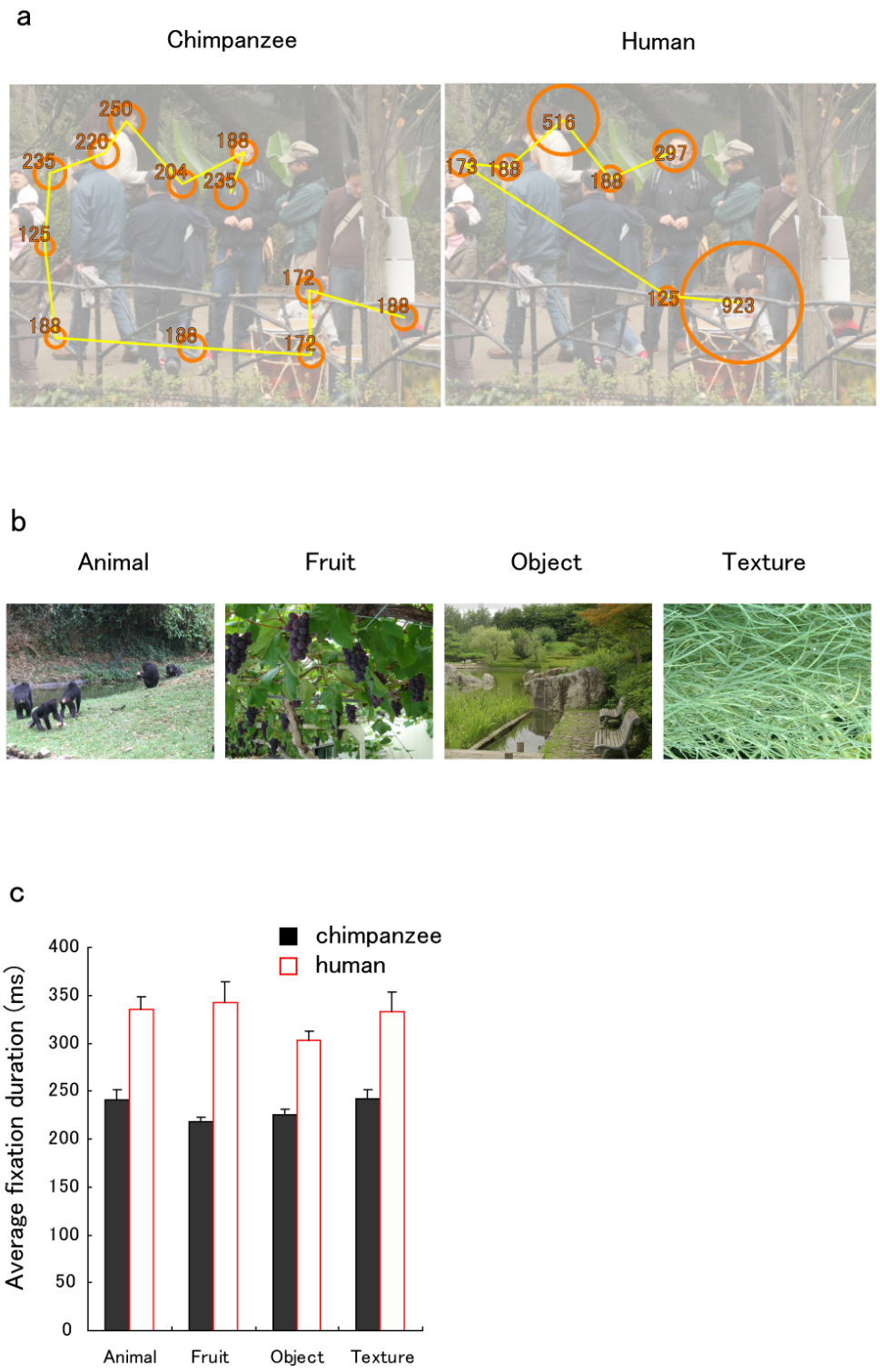


Figure 2-3-1

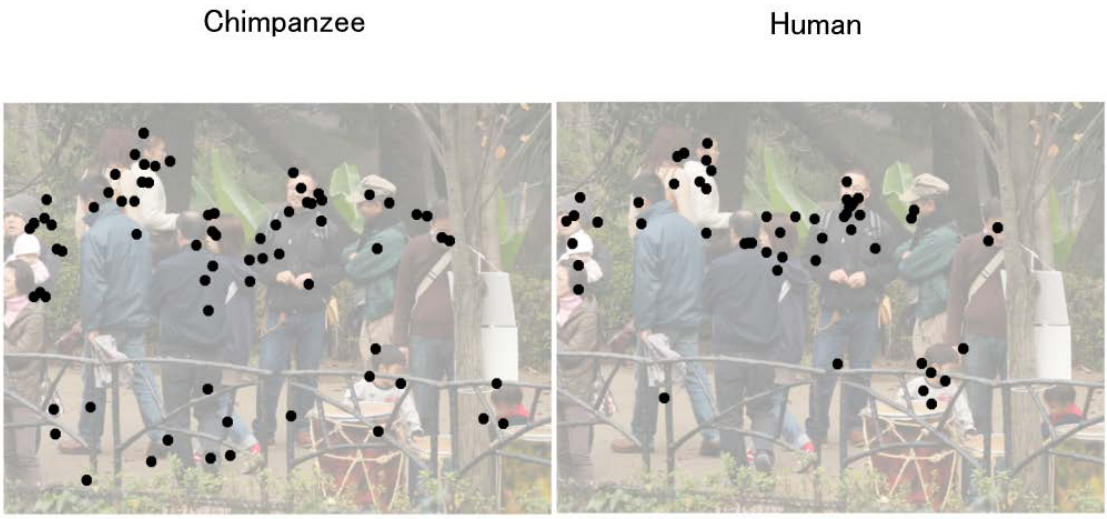


Figure 2-3-2

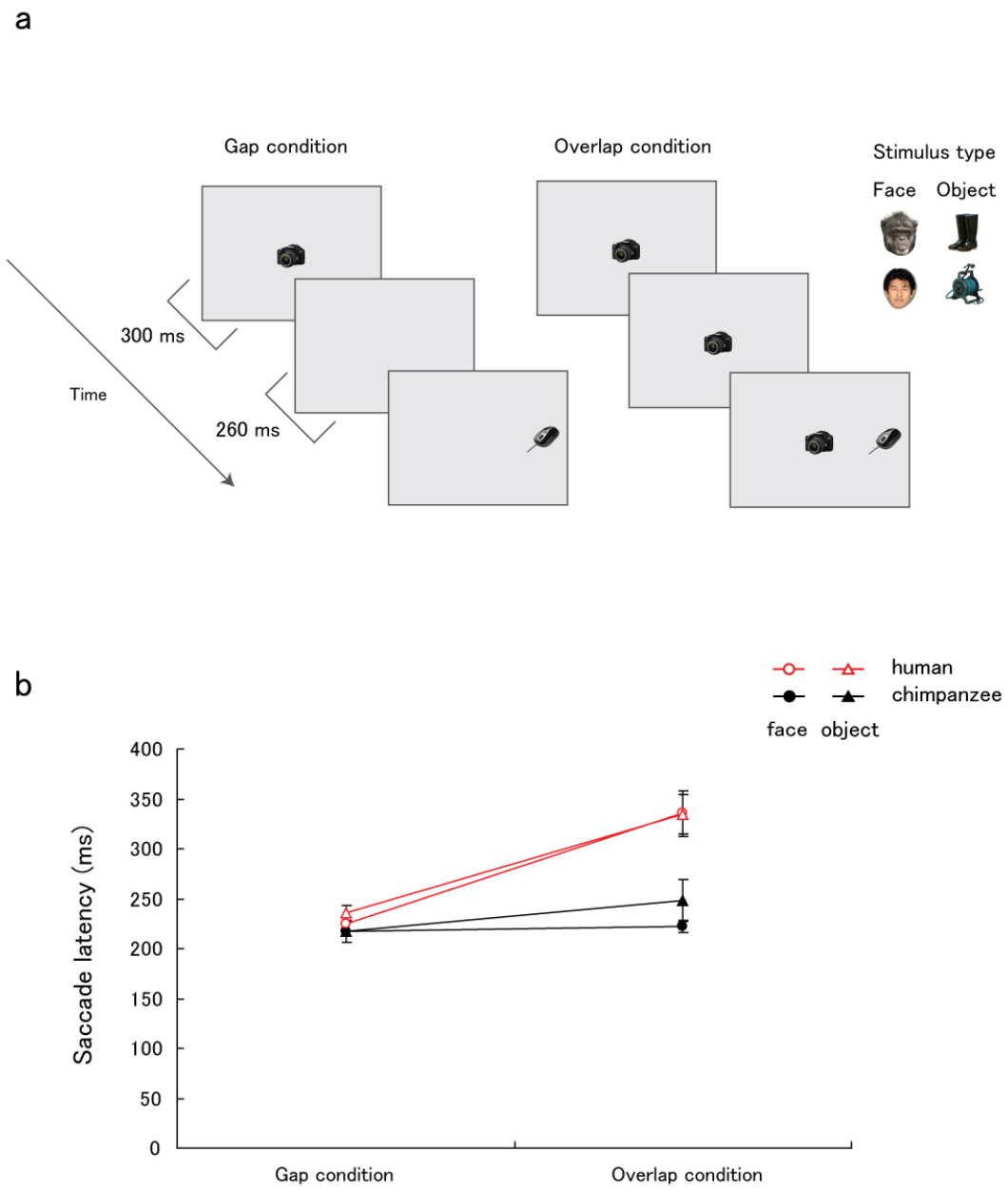

a
Gap condition
Overlap condition
Stimulus type
Face
Object
300 ms
Time
260 ms
b
human
chimpanzee
face
object
Saccade latency (ms)
400
350
300
250
200
150
100
50
0
Gap condition
Overlap condition

Figure 2-3-3

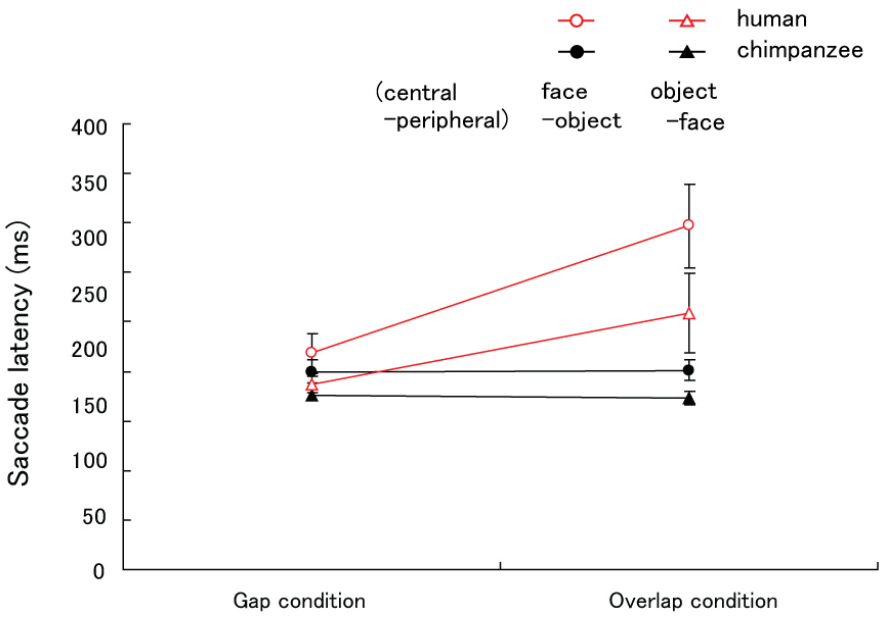

human
chimpanzee
(central
−peripheral)
face
−object
object
−face
Saccade latency (ms)
400
350
300
250
200
150
100
50
0
Gap condition
Overlap condition

Figure 2-3-4

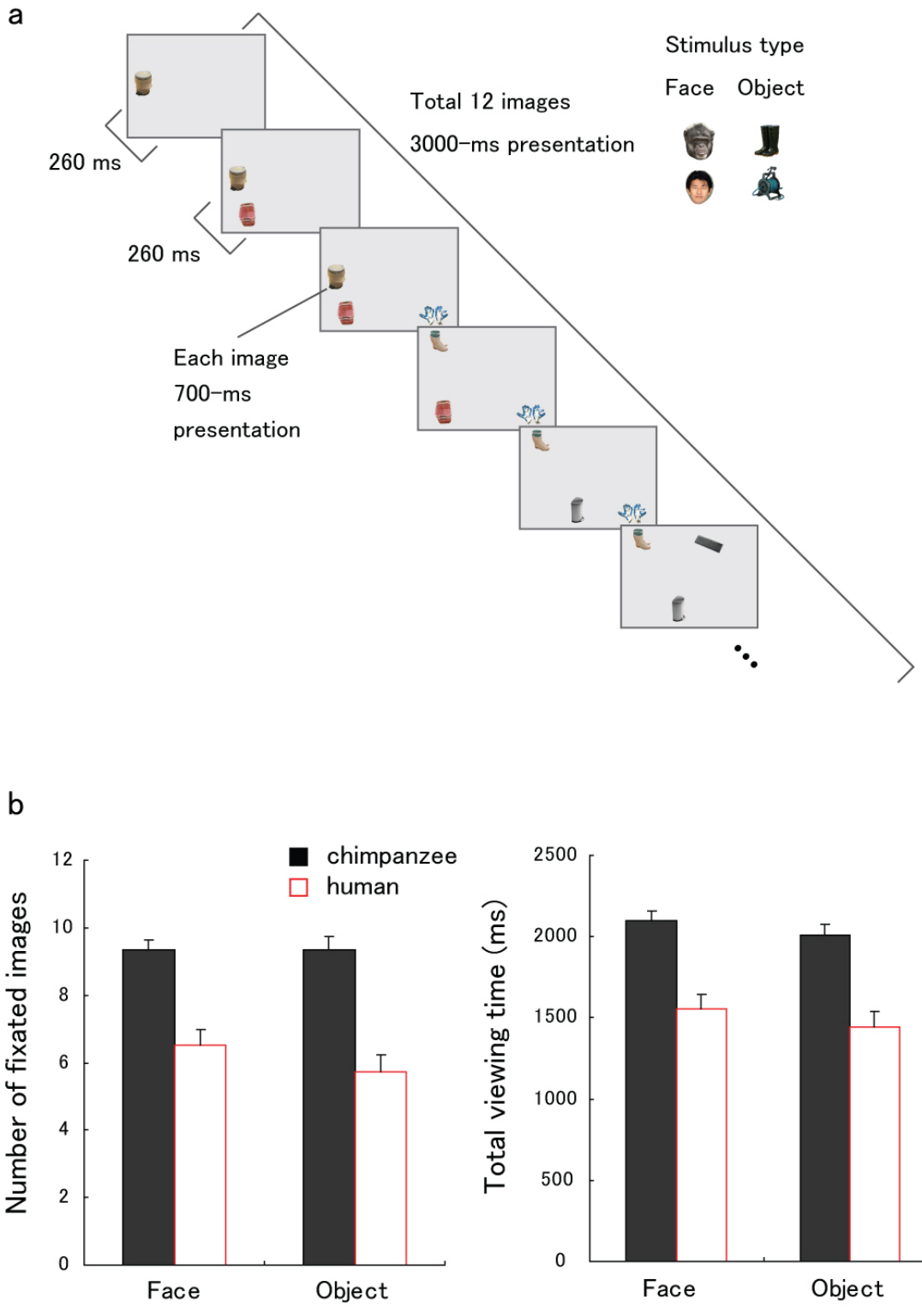

a
Stimulus type
Face Object
Total 12 images
3000–ms presentation
260 ms
260 ms
Each image
700–ms
presentation
b
chimpanzee
human
Number of fixated images
0
2
4
6
8
10
12
Face
Object
Total viewing time (ms)
0
500
1000
1500
2000
2500
Face
Object

Figure 2-3-5

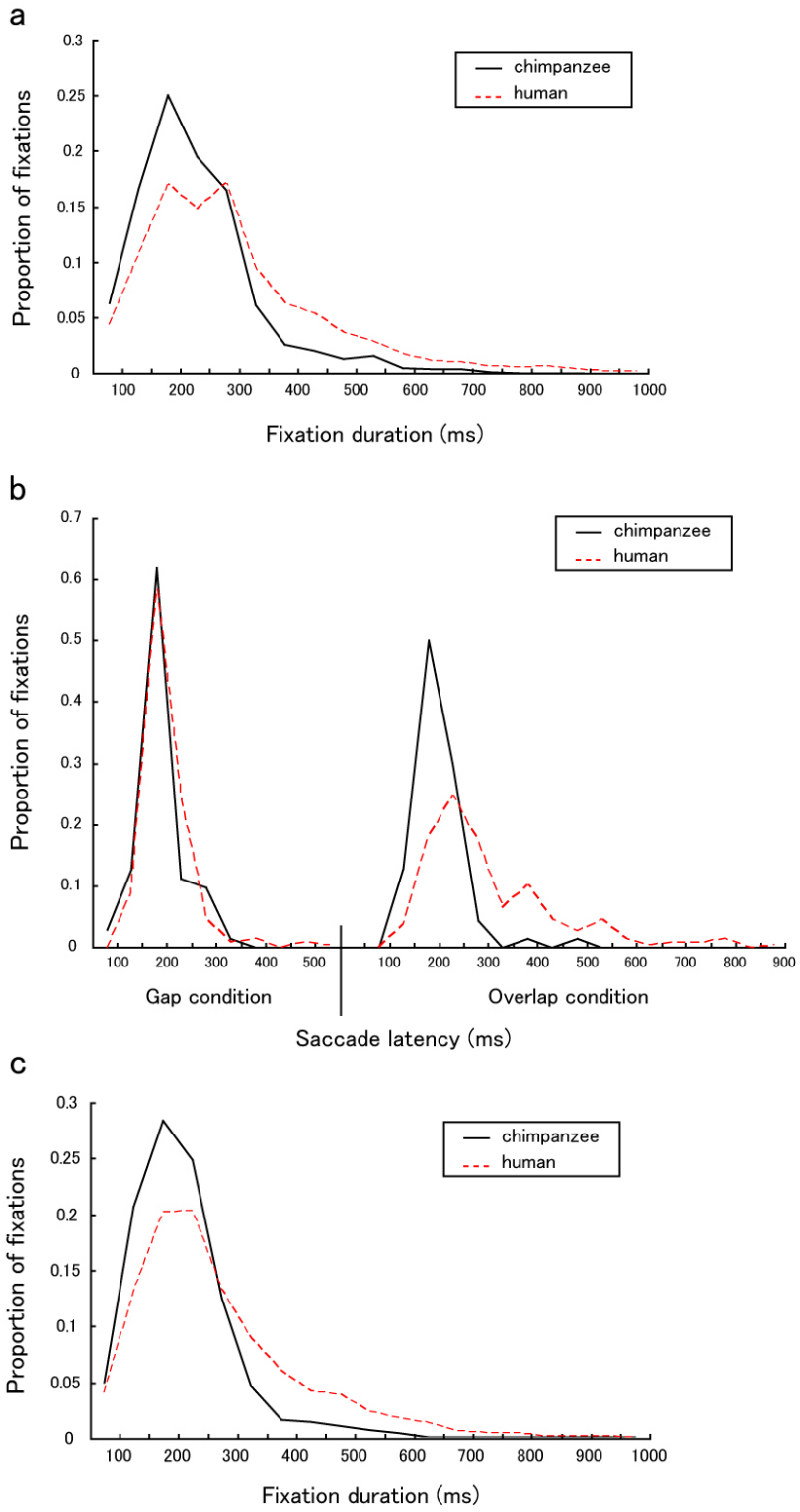

a
Proportion of fixations
chimpanzee
human
Fixation duration (ms)
b
Proportion of fixations
chimpanzee
human
Gap condition
Overlap condition
Saccade latency (ms)
c
Proportion of fixations
chimpanzee
human
Fixation duration (ms)

Figure 2-3-6

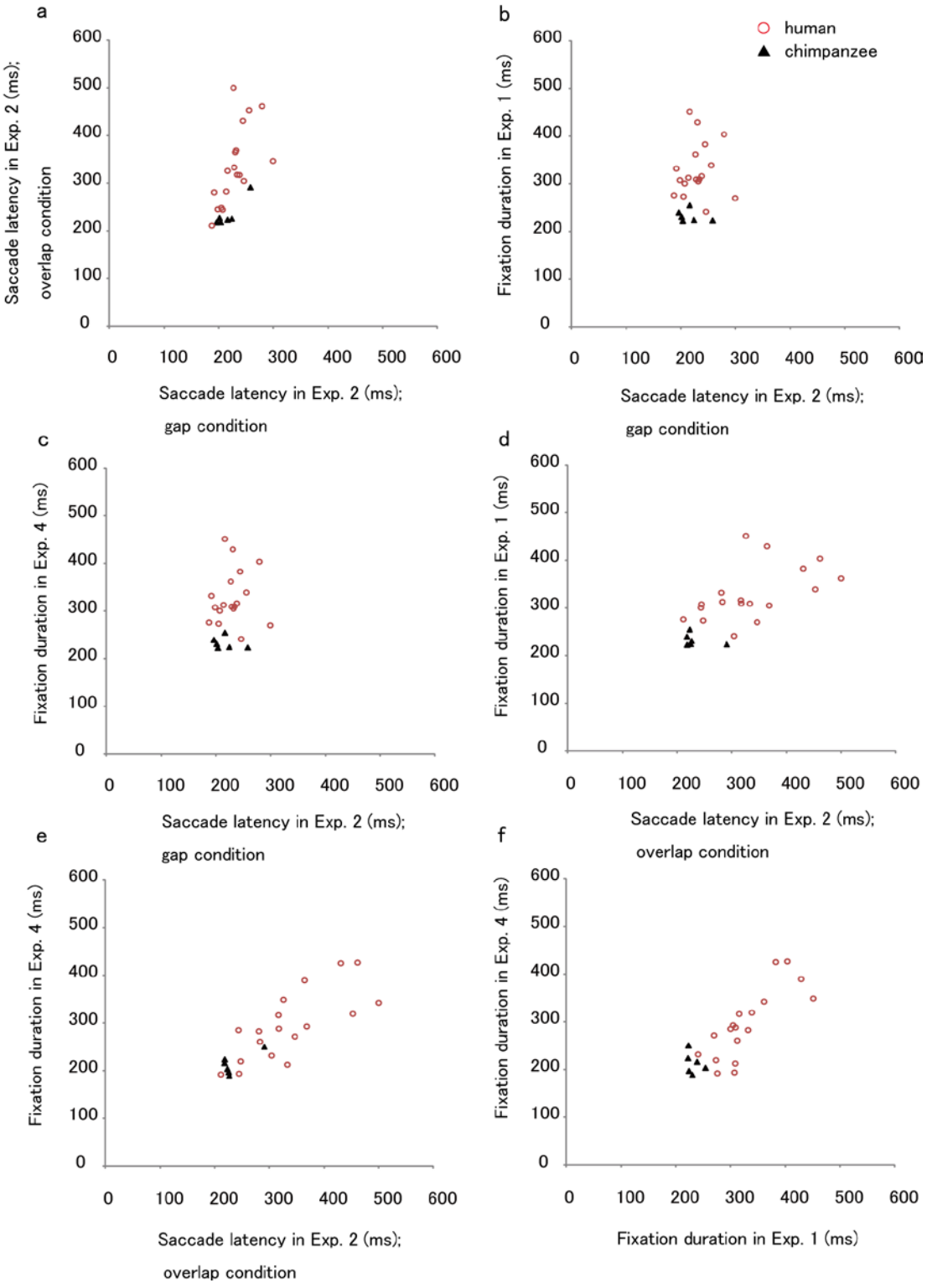

a
Saccade latency in Exp. 2 (ms); overlap condition
Saccade latency in Exp. 2 (ms); gap condition
b
human
chimpanzee
Fixation duration in Exp. 1 (ms)
Saccade latency in Exp. 2 (ms); gap condition
c
Fixation duration in Exp. 4 (ms)
Saccade latency in Exp. 2 (ms); gap condition
d
Fixation duration in Exp. 1 (ms)
Saccade latency in Exp. 2 (ms); overlap condition
e
Fixation duration in Exp. 4 (ms)
Saccade latency in Exp. 2 (ms); overlap condition
f
Fixation duration in Exp. 4 (ms)
Fixation duration in Exp. 1 (ms)
0
100
200
300
400
500
600

Figure 2-3-7

Chapter 2-4.

Face scanning in chimpanzees and humans: continuity and discontinuity

Introduction

Humans have highly sophisticated forms of facial communication. Faces and eye movements contribute to a variety of expressions that are then efficiently perceived in humans. Moreover, humans often engage in lengthy face-to-face communications, accompanied by intense eye contact (Argyle & Cook, 1976). How and when such unique forms of human communications evolved have long been of interest, particularly from a comparative perspective that addresses how and to what extent human and non-human animals are similar and dissimilar in their forms of communications (Chevalier-Skolnikoff, 1973; Darwin, 1999). Chimpanzees (*Pan troglodytes*), the species closest to humans, are known to share several forms of facial communications with humans. Field-observational, morphological, and experimental studies have identified similarities between the species. For example, both frequently engage in mutual gazing during mother-offspring interactions (Bard et al., 2005; van Lawick-Goodall, 1967). Additionally, the facial musculature of chimpanzees is homologous with that of humans in several respects and both species exhibit some similar facial expressions (e.g., play faces: Chevalier-Skolnikoff, 1973; Parr, Waller, Vick, & Bard, 2007; Vick, Waller, Parr, Pasqualini, & Bard, 2007). Experimental studies have shown that the mechanisms underpinning face/gaze perception in chimpanzees are similar to those in humans, including the ability to follow the gaze of another (Itakura & Tanaka, 1998; Okamoto et al., 2002; Tomasello, Hare, Lehmann, & Call, 2007) and the holistic (configural) processing of faces (Parr, Dove, & Hopkins, 1998; Tomonaga, 1999, 2007b); while the evidence for holistic face processing remains controversial in monkeys (Adachi, Chou, & Hampton, 2010; Dahl, Wallraven, Bulthoff, & Logothetis, 2009; Parr & Heintz, 2008; Tomonaga, 1994).

On the other hand, notable differences between humans and chimpanzees have also been identified. Face-to-face communications in chimpanzees tend to occur in brief spurts and during arousing situations, and the most typical close-range long-bout affiliative communications are tactile

(e.g., grooming: Goodall, 1968; van Hooff, 1973). Humans have fine motor control of the muscles around the eyes, enabling the formation of subtle expressions in these regions (Ekman & Friesen, 1978). In contrast, these regions are hardly visible in chimpanzees because of their high eyebrow ridges and dark eyes (Kobayashi & Kohshima, 1997, 2001; Parr et al., 2007). Instead, chimpanzees have fine motor control of their lips, enabling the formation of rich expressions in the mouth region (Chevalier-Skolnikoff, 1973; Parr et al., 2007; Vick et al., 2007). Kobayashi and Kohshima (2001) found that, compared with other primates, humans have exceptionally large white sclera that clearly contrast with the colours of their iris and skin. These authors hypothesized that human eyes have adapted to enhance gaze signals, such as eye directions. Experimental studies have shown that chimpanzees were less sensitive than human infants to the eye directions of human experimenters when the experimenter's head was immobile (Tomasello et al., 2007), possibly because chimpanzees inspected the experimenter's face more briefly during their interactions (Carpenter & Tomasello, 1995). Additionally, in an experiment using a computerized task, chimpanzees did not show overt evidence of reflexive shifts of attention in the same direction as the eyes in a human stimulus face (Tomonaga, 2007a).

These results consistently suggested that eye regions (upper faces) are especially important in humans, compared with chimpanzees, in facial communications. Fine muscles and eye movements are accentuated by hair and colour-contrast in the eye regions in humans, signalling a variety of communicative intents. These morphologies are visually salient and thus might independently attract the attention of other individuals. However, given that chimpanzees were less sensitive than were humans to the eye direction of humans, humans might also have characteristic behavioural/attentional sensitivities to the eye regions. That is, humans might have stronger behavioural tendencies than chimpanzees to actively collect information from the eye regions.

Carpenter et al. (1995) found that, in interactive situations (facilitative of joint attention),

human infants looked at the experimenter's face an average of twice as long as chimpanzees. Interestingly, in this related study, the human-raised and enculturated chimpanzees did not differ from those reared by biological mothers in this regard. Kano and Tomonaga (2009) used the eye-tracking method to measure how humans (adults) and chimpanzees viewed whole-body images of conspecifics and non-conspecifics. Both species showed highly similar patterns of scanning and both paid more attention to the face than to the other body regions depicted in the images. Additionally, both species showed, on average, an equal number of fixations on the faces. The critical difference was that the average durations of fixation on the faces were shorter in chimpanzees (300 ms) than in humans (680 ms). These differences were consistent across conspecific and non-conspecific images. However, the patterns of face scanning, the details of eye gaze to facial features (e.g., eyes, mouth) remain unclear because eye gaze was disproportionately directed to faces versus the rest of the body in these previous studies.

This study thus aimed to determine the patterns of face scanning in chimpanzees. We examined the spatial (where) and temporal (when) characteristics of fixation sequences, and compared them with humans. We presented close-up shots of faces to enhance our ability to observe eye movements as participants scanned the images and measured frequencies, durations and probabilities of fixations. Although widely employed in humans (Henderson, Williams, & Falk, 2005; Pelphrey et al., 2002; Walker-Smith, Gale, & Findlay, 1977; Yarbus, 1967) and monkeys (Gothard, Erickson, & Amaral, 2004; Guo, Robertson, Mahmoodi, Tadmor, & Young, 2003; Keating & Keating, 1982; Mendelson, Haith, & Goldmanrakic, 1982; Nahm, Perret, Amaral, & Albright, 1997; Sato & Nakamura, 2001), this is the first study investigating face scanning in apes. Humans and monkeys are known to intensely scan the main features of faces (i.e., eyes, nose, and mouth). It is also known that these species look at the eye regions longer than at other facial features and are more likely to initially inspect the eye regions. The comparisons between humans and monkeys have

thus suggested qualitative similarities in face scanning. However, the direct, quantitative comparison of face scanning between human and nonhuman primates is practically lacking partly because of the difference in the experimental procedures and of the phylogenetical distance between humans and monkeys. This study directly compared humans and their closest evolutionary relatives, chimpanzees, using non-invasive eye-tracking method under unrestrained conditions, and thereby aimed to assess both qualitative and quantitative similarities and differences between the species. The fixation sequences of humans were characterized by sequential fixations over the eyes and mouth, which appeared to somewhat resemble inverted triangular traces (Walker-Smith et al., 1977; Yarbus, 1967). The precise comparisons between species enabled us to analyze the characteristic patterns underpinning the sequential/spatiotemporal aspects of eye movements in each species.

We conducted the two experiments. Experiment 1 presented both conspecific and non-conspecific faces and examined general similarities and differences in face scanning between chimpanzees and humans. We have addressed three questions in experiment 1. The first question addressed if the patterns of scanning in chimpanzees and/or humans depend on the specific responses to face stimuli; that is, if they do not solely depend on the general responses to relatively broad class of stimuli, or if they do not solely depend on the low-level guidance of eye movements caused by visual salience of image features (e.g. white sclera in humans, protruded nose in humans, protruded mouth in chimpanzees). To confirm these issues, we examined the patterns of scanning especially for the main features of faces (i.e., eyes nose, mouth) because the previous studies about monkeys and humans suggested that the patterns of scanning specific to face stimuli were characterized by the systematic responses to those main features. We also examined the consistency in viewing patterns for faces between the current and previous (Kano & Tomonaga, 2009) studies which presented with the close-up shot of faces and full shot of whole bodies, respectively, in order to examine the effect of global view (i.e. image scale) on the eye movements. The second question

addressed what the characteristic patterns of face scanning in chimpanzees and humans are. Given the previous studies mentioned earlier, we expected that chimpanzees and humans would differ especially with regard to viewing patterns for the eye regions. The third question addressed on what natures of faces those characteristic patterns of face scanning in chimpanzees and humans depend. As known in previous studies, their characteristic patterns of face scanning might be influenced by the experience of exposure to certain types of faces (e.g., own/other race: Michel, Rossion, Han, Chung, & Caldara, 2006; reared experience by own/other species; Martin-Malivel & Okada, 2007) and by the phylogenetical relatedness (e.g. conspecific/non-conspecific: Pascalis & Bachevalier, 1998). Their characteristic patterns might reflect more general responses to the limited sets of stimulus cues, such as face-like configurations. Experiment 2 presented to chimpanzees and humans their species-specific facial expressions. Face-to-face interactions typically involve various emotional gestures in both species (Argyle & Cook, 1976; van Hooff, 1967). Although several direct comparisons of gaze perception in humans and chimpanzees have been conducted, as mentioned earlier, such attempts have not addressed the perception of facial expressions. Experiment 2 further examined the questions addressed in experiment 1. The advanced question of experiment 2 addressed how the characteristic patterns of face scanning by each species change according to the type of expressions. Because the quality of information in eye and mouth regions somewhat differ between chimpanzee and human expressions as mentioned earlier, we expected that the species differences would be characterized by their respective responsiveness to eye and mouth regions of each expression.

Methods

We used an identical experimental framework, with the exception of the stimuli, as those used by Kano & Tomonaga (2009) (these are referred to as 'the previous experiments" in the Methods section) to allow comparisons between the studies.

Experiment 1

Participants and apparatus. Six chimpanzees (5 females, 1 male; aged 8-31) and 18 humans (11 females, 7 males; aged 18-31; all Japanese students) participated in Experiment 1. All chimpanzees and half the humans had participated in the previous experiments. The chimpanzees were members of a social group comprised of 14 individuals living in an enriched environment with 700-m2 outdoor compound and an attached indoor residence (Matsuzawa, Tomonaga, & Tanaka, 2006). The outdoor compound was equipped with 15-m-high climbing frames, small streams, and various species of trees (Ochiai & Matsuzawa, 1997). Access to the outdoor compound was given for each individual every other day and in the daytime. Daily meals included a wide variety of fresh fruits and vegetables fed throughout the day, supplemented with nutritionally-balanced biscuits (fed twice daily), and water available ad libitum. Both chimpanzee and human participants had extensive experience interacting with both species and were thus highly familiar with both kinds of faces. The chimpanzees, like the human participants, had extensive experience observing photographs of faces (Matsuzawa et al., 2006) and thus never responded fearfully to the faces. No food or water deprivation was conducted during the study period. Care and use of the chimpanzees adhered to the 2002 version of the Guideline for Care and Use of Laboratory Primates by the Primate Research Institute, Kyoto University. The experimental protocol was approved by the Animal Welfare and Care Committee of the institute. Informed consent was obtained from all human participants. Both species used the same apparatus to allow for direct comparisons. Participants sat still and unrestrained in an experimental booth and viewed a 17-inch LCD display (1280 × 1024 pixels) at a

distance of approximately 60 cm. A table-mounted eye-tracker measured their eye movements using infrared corneal reflection techniques (60 Hz; Tobii X120, Tobii Technology AB; Appendix 1). As a result of the training conducted during previous experiment, chimpanzees were already skilled at sitting still in front of an eye-tracker and looking at a fixation point that appeared on the screen. We conducted two-point and five-point calibrations for chimpanzees and humans, respectively. Calibrations were repeated until maximum accuracy was obtained, resulting in high spatial resolution for the eye tracking in both species. In the preliminary recording, six participants of each species were asked to look at a fixation point and the error values—the average distance between the intended and the recorded fixations—were measured as 0.62±0.06 of a degree (mean±SEM) for the chimpanzees and 0.52±0.05 of a degree for the humans. Refer to the previous experiment for details about calibration training and procedures.

Stimuli and procedures. Stimuli consisted of 24 colour still photographs of the faces of 17 species of non-primate mammals (giraffes, rhinos, lions, etc.; hereafter, mammals), 24 photographs of chimpanzee faces and 24 photographs of human faces (frontal views; 12 individuals, six of whom were familiar and six of whom were unfamiliar to participants; see Figure 2-4-1 for examples). Both species had relatively less experience with observing mammal than chimpanzee and human faces. Half of these facial images were extracted from the whole-body images used in the previous experiments. The photographs were converted into 1000 × 800 pixels with surrounding gray frames (1280 × 1024 pixels in total). During testing, each trial was initiated by the participants looking at a fixation point that appeared at a random position on the screen. The photographs were then presented for 2 s, and participants were allowed to move their eyes freely to view the photograph. A total of 72 photographs were presented to the humans within a single day, whereas the sessions were divided among 10 days for the chimpanzees to maintain their spontaneous motivation for viewing photographs. The presentation order of photographs was randomized within the entire sessions

respectively for each participant. The human participants received 500 yen after the session, and the chimpanzees obtained a small piece of apple after each trial, regardless of their viewing behaviours.

Figure 2-4-1

Experiment 2

Five chimpanzees (4 females, 1 male) and 9 humans (7 females, 2 males) participated in Experiment 2. One chimpanzee (adult female) was eliminated from testing because of her lack of attention to the stimuli. The stimuli consisted of 12 colour still photographs portraying species-specific chimpanzee facial expressions (three each portraying neutral faces, hoot faces, scream faces, and compressed-lip faces; see Parr et al., 2007 for descriptions of these expressions) and 12 monochrome still photographs of a standardized set of human facial expressions (three each portraying neutral faces, happy faces, fearful faces, and angry faces; taken from Ekman & Friesen, 1978; see Figure 2-4-3 for examples). Chimpanzee facial expressions were obtained by videotaping social interactions at another chimpanzee colony and then isolating the frames containing the expressions with peak intensity. The photographs of chimpanzee and human expressions were converted into 1000 × 800 pixels and 600×800 pixels, respectively (no background in human photographs), with the surrounding gray frames (1280 × 1024 pixels in total). Facial expressions were presented for 2 s to chimpanzees. Given the slower inspection of each facial feature by humans (shorter fixation durations on each feature; see below) than by chimpanzees, which was found in the experiment 1, faces were presented for 5 s to humans in order to leave sufficient time for them to explore each facial feature. Each trial presented a photograph that was randomly drawn from each type of facial expression. The remaining procedures were the same as those used in Experiment 1.

Data Analysis

Trials in which participants did not view the monitor for more than 300 ms were eliminated from the analysis, resulting in a loss of 4.1% of the data obtained from chimpanzees (no data were lost for humans). During Experiment 2, we repeated these trials (8.3% of all the data obtained from chimpanzees) after the end of session, resulting in no loss of data for either species. To define areas of interest (AOI), the scenes were divided into the face regions and the rest of the scenes. The face regions were further divided into the eye, nose (or referred as 'mid-face' in mammal faces), mouth, and other regions (see Figure 2-4-1 for examples). To avoid errors in gaze estimations, AOIs were drawn slightly larger than the actual outlines (approximately 20 pixels on the edges). A fixation was scored if the gaze remained stationary (within a radius of 50 pixels) for at least 75 ms (more than 5 measurement samples). Otherwise, the recorded sample was defined as part of a saccade. We excluded the samples recorded during the first 200 ms, thereby eliminating fixations that followed the offset of the fixation spot. We used four dependent variables in this study: cumulative viewing time, number (frequency) of fixations, average fixation duration, and the probability of fixations (as a function of fixation order). The probability of fixations was calculated as the proportion of photographs in which a certain AOI was the target at a certain fixation order. In experiment 2, the number of fixations was calculated as a proportion of the total numbers of face fixations to correct for the differences in presentation duration between the two subject species. For the statistical analyses, we distinguished between within-species and between-species comparisons. For within-species comparisons, we tested for the difference in viewing patterns for each AOI within each subject species. For between-species comparisons, we tested for the interaction between subject species and AOIs in viewing patterns. In ANOVAs, subject species and other independent variables (facial features, fixation order, facial expressions, and presentation sections) served as between- and within-subject factors, respectively. In cases in which the assumption of homogeneity of variance was violated, the Greenhouse-Geisser correction was applied, and corrected p values were calculated.

Post-hoc tests for within- and between-species comparison used t-tests, and Bonferroni's corrections corrected the alpha level (set at 0.05) for the number of comparisons. We conducted all analyses independently for each stimulus species (i.e., chimpanzees, humans, and mammals, and thus did not include stimulus species as a factor for ANOVA because facial morphologies (the proportion of each facial feature) differed somewhat according to stimulus species. In experiment 2, to compare between facial expressions in viewing patterns, we normalized the data for each expressive face according to the proportions characterizing the neutral faces of the same stimulus species. We then compared the data between facial expressions independently for each stimulus species.

Results

Experiment 1

We obtained several results consistent with those of previous experiments (Kano & Tomonaga, 2009) involving the presentation of the whole-body images of the same animals. First, both species fixated on the face region more frequently than on the rest of the scenes for all species (chimpanzee faces: 4.5 vs. 1.3 times, 4.2 vs. 0.14 times; human faces: 3.8 vs. 2.0 times, 3.7 vs. 0.23 times; mammal faces: 4.6 vs. 1.5 times, 4.1 vs. 0.23 times, for chimpanzee and human participants respectively; *ps* < 0.05). Second, chimpanzees and humans did not differ in the numbers of fixations on the face region for all species (*ps* > 0.05). Third, the average fixation durations on the faces were shorter in chimpanzees than in humans (see Table 2-4-1).

Table 2-4-1

The new findings emerging from this study involved similarities and differences between the species with regard to the scanning patterns for each facial feature. We focussed on the number of fixations and on the probability of fixations as a function of fixation order (i.e., eye movement path; the variables represented by the yellow lines in Figure 2-4-1) in deriving the following results because we obtained similar results when the other two variables were included in the analysis (i.e., cumulative viewing time and average fixation duration; the variables indicated by the orange circles in Figure 2-4-1; see Table 2-4-1 for these results). Chimpanzees and humans fixated on the main facial features (i.e., eyes, nose, and mouth) more frequently than they fixated on the other regions for chimpanzee faces (Figure 2-4-2a; $t_5 = 5.27$, $p = 0.003$, $t_{17} = 24.40$, $p < 0.001$, respectively) and human faces ($t_5 = 6.70$, $p = 0.001$, $t_{17} = 17.71$, $p < 0.001$, respectively). Although this pattern of results did not emerge with regard to mammal faces in chimpanzees ($t_5 = 1.51$, $p = 0.19$; but

emerged in humans; $t_{17} = 8.77$, $p < 0.001$), this probably is simply due to the frequent inspection of the other regions (e.g. a mane of lion, a hone of rhino) both in chimpanzees and humans. The ANOVA focussing on the main facial features revealed a significant interaction between subject species and facial features (chimpanzee faces: $F_{1.1, 24} = 11.5$, $p < 0.001$, $\eta^2 = 0.34$; human faces: $F_{1.2, 26} = 6.03$, $p = 0.005$, $\eta^2 = 0.21$; mammal faces: $F_{1, 22} = 44.7$, $p < 0.001$, $\eta^2 = 0.67$). Post-hoc comparisons between the species revealed that humans fixated on the eye region more frequently than did chimpanzees (Figure 2-4-2a; chimpanzee faces: $t_{22} = 2.93$, $p = 0.008$; human faces: $t_{22} = 2.73$, $p = 0.012$; mammal faces: $t_{22} = 5.30$, $p < 0.001$) and chimpanzees fixated on the mouth region more frequently than did humans (chimpanzee faces: $t_{22} = 4.21$, $p < 0.001$; human faces: $t_{22} = 3.34$, $p = 0.003$; mammal faces: $t_{22} = 3.62$, $p = 0.002$). Post-hoc comparisons within the species revealed that, unlike the humans (chimpanzee faces: $t_{17} = 9.29$, $p < 0.001$; human faces: $t_{17} = 7.79$, $p < 0.001$; mammal faces: $t_{17} = 13.9$, $p < 0.001$), the chimpanzees did not fixate on the eye regions more frequently than they fixated on the mouth regions (chimpanzee faces: $t_{17} = 1.44$, $p = 0.20$; human faces: $t_{17} = 3.92$, $p = 0.011$; mammal faces: $t_{17} = 1.26$, $p = 0.26$).

Figure 2-4-2

Figure 2-4-2b illustrates the temporal character of the aforementioned differences between the species for the eyes and mouth, respectively. An ANOVA revealed significant interactions between subject species and fixation order for the eyes (chimpanzee faces: $F_{3, 66} = 3.63$, $p = 0.017$; human faces: $F_{2.0, 45} = 4.36$, $p = 0.007$, $\eta^2 = 0.16$; mammal faces: $F_{3, 66} = 11.68$, $p = 0.001$, $\eta^2 = 0.34$) and for the mouth (chimpanzee faces: $F_{3, 66} = 3.82$, $p = 0.014$, $\eta^2 = 0.14$; human faces: $F_{3, 66} = 4.26$, $p = 0.008$, $\eta^2 = 0.16$; mammal faces: $F_{3, 66} = 9.13$, $p < 0.001$, $\eta^2 = 0.29$). We then conducted post-hoc tests for each fixation order. Both species scanned the eyes, followed by the mouth, as indicated by

the higher probability of first fixations on the eye region than on the mouth region (*ps* < 0.05 for both species viewing all species of faces). Additionally, the probabilities for first fixations did not differ significantly between the species for the eyes or mouth (*ps* > 0.05). However, chimpanzees were less likely than were humans to fixate on the eye region during later fixations (*ps* < 0.01). Chimpanzees were more likely than humans to fixate on the mouth region as their second fixation (*ps* < 0.01), but not as their third or fourth fixation (*ps* > 0.05). That is, although chimpanzees and humans both began their fixation sequences with the eye regions, humans were more likely than chimpanzees to subsequently re-fixate on the eye regions, whereas the chimpanzees were more likely than humans to subsequently shift their gazes to the mouth region.

We should rule out two possible effects of the presentation procedures on the results. First, the chimpanzees might have habituated to the faces more quickly than did the humans, enabling them to scan the faces more rapidly. We divided the entire sessions into three sections and compared the number of fixations on the face regions among these three sections. However, an ANOVA did not find any significant interaction between subject species and section (chimpanzee faces: $F_{2, 44} = 1.62$, $p = 0.20$, $\eta^2 = 0.069$; human faces: $F_{2, 44} = 0.376$, $p = 0.19$, $\eta^2 = 0.072$; mammal faces: $F_{2, 44} = 1.96$, $p = 0.15$, $\eta^2 = 0.082$). Second, the presentation duration in this study (2 s) might have been too short for the humans, leaving insufficient time to explore facial features other than eyes. Thus, the same five human participants viewed half of the same human and chimpanzee faces again, but with a presentation duration of 10 s. We found that the moderate decrease/increase in the probability of fixation on the eye/mouth regions by humans (as seen in Figure 2-4-2b) remained during this longer presentation and that the decrease/increase stabilized at the seventh fixation at a probability of 0.4 0.6 for the eye region and of 0.0 0.2 for the mouth region. Thus, we confirmed that the aforementioned differences between species were not related to presentation order or presentation duration.

Experiment 2

Figure 2-4-3 illustrates the attentional responses of chimpanzees and humans to species-specific facial expressions of both species. Only eyes and mouth regions of the chimpanzee and human faces (mobile regions in facial expressions) were included in this analysis to avoid redundancy. We first conducted ANOVAs (subject species × facial features × facial expressions) and found the significant interaction between the three factors (chimpanzee faces: $F_{3, 36} = 4.14$, $p = 0.013$, $\eta^2 = 0.25$; human faces: $F_{3, 36} = 1.62$, $p = 0.007$, $\eta^2 = 0.28$). We then conducted ANOVAs (subject species × facial expressions) respectively for eye and mouth regions and found that humans viewed the eye regions more frequently than did chimpanzees (chimpanzee faces: $F_{1, 12} = 8.41$, $p < 0.001$, $\eta^2 = 0.89$; human faces: $F_{1, 12} = 15.2$, $p < 0.001$, $\eta^2 = 0.94$), and chimpanzees viewed the mouth regions more frequently than did humans (chimpanzee faces: $F_{1, 12} = 40.1$, $p < 0.001$, $\eta^2 = 0.96$; human faces: $F_{1, 12} = 17.1$, $p < 0.001$, $\eta^2 = 0.92$). ANOVAs (facial features × facial expressions) revealed that, while chimpanzees differentiated between the facial expressions in terms of the proportion of fixations to eye or mouth regions (chimpanzee faces: $F_{3, 12} = 8.37$, $p = 0.003$, $\eta^2 = 0.67$; human faces: $F_{3, 12} = 11.0$, $p = 0.001$, $\eta^2 = 0.73$), humans did not (chimpanzee faces: $F_{1.3, 11} = 0.61$, $p = 0.615$, $\eta^2 = 0.071$; human faces: $F_{3, 24} = 0.29$, $p = 0.82$, $\eta^2 = 0.036$). To examine the effect of different presentation duration in humans (5 s) and chimpanzees (2 s), we conducted the same analyses for the first 2 s of presentation in humans. However, we confirmed the similar tendency in humans (the intense and persistent eye-viewing) in this analysis. These results indicate that, while chimpanzees changed their scanning patterns in response to the facial actions, humans maintained their intense focus on the eye regions across expressions. This difference between the species was consistent across chimpanzee and human faces.

Figure 2-4-3

Discussion

To our knowledge, this is the first reported study to obtain comparative data on face scanning in chimpanzees and humans, thereby offering methodological advances for examining the evolution of face perception and facial communication. The observed scanning patterns were not dependent solely on the visually salient features in the faces, as evidenced by the systematic similarities and differences between the species in the scanning patterns for main features of faces. We confirmed the robust consistency between the current and previous study (Kano & Tomonaga, 2009) involving the presentation of close-up shots of faces and full shots of whole bodies, respectively. These results indicated that chimpanzees and humans elicited the patterns of scanning specific to face stimuli. Several qualitative similarities highlighted the homologous nature of face scanning between chimpanzees and humans: both demonstrated intense scanning of the main facial features (except in mammal faces; see Results) and the same order of inspection for each facial feature (from the eyes to mouth). Importantly, notable quantitative differences between chimpanzees and humans also emerged: the prolonged eye-viewing by humans (the sequential re-fixations on the eyes), the quick scanning of faces by chimpanzees (immediate shifts of eye gaze from the eyes to mouth), and the persistent preferences among humans for eyes across various facial expressions. If we emphasized the differences in the inverted triangular sequences of fixations over the eyes and mouth exhibited by humans, the sequences of fixations exhibited by chimpanzees would appear to resemble vertical segments of lines (as seen in Figure 2-4-1).

These similarities and differences between the species were consistent across conspecific faces, non-conspecific faces, and even phylogenetically distant mammal faces. Recall that the chimpanzees and humans in this study were highly familiar with chimpanzee and human faces, but relatively unfamiliar with the mammal faces (see Methods). Those patterns thus did not appear to reflect the experiences of exposures to a certain type of faces or the phylogenetical relatedness, but

rather involved more general responses to the face-like configurations. In experiment 2, we found that chimpanzees and humans viewed the mouth and eye regions more frequently than did the other species, respectively, consistently with experiment 1. Additional findings of experiment 2 involved that, while chimpanzees changed their scanning patterns in response to those facial actions, whereas humans maintained intense eye-viewing across the expressions. These species differences were consistent across chimpanzee and human faces, which was also consistent with experiment 1.

Overall, the patterns of face scanning in humans were characterized as the prolonged eye-viewing and its persistency across facial expressions, while those in chimpanzees were characterized as the quick, vertical scanning for faces, the frequent inspection of the mouth regions, and the responsiveness to the facial actions. The characteristics in humans indicate their active viewing for the eye regions, which is independent of various facial morphologies and expressions. On the other hand, although the overall patterns of eye movements in chimpanzees were characterized as the specific responses to face stimuli (see above), some of the characteristics may be attributable to either general responses to relatively broad class of stimuli or low-level guidance of eye movements caused by visual salience of image features. For example, the quick scanning in chimpanzees may reflect their general ability of scanning for scenes. In the previous experiment (Kano & Tomonaga, 2009), chimpanzees elicited the quicker shift of fixation location than did humans for overall scenes (i.e., not only for faces but also for the rest of the scenes). Additionally, the frequent inspection of mouth regions in chimpanzees may be the passive viewing for the visually salient regions rather than the active viewing for the informative regions, reflecting their reluctance of eye-viewing. The mouth regions are not only informative for emotions but also visually salient especially in chimpanzee faces and their expressions. Thus, although both species have the high motivation to view faces (but to the less extent in chimpanzees), the attention may be less focused and more distributed in space in chimpanzees when viewing facial features.

The salient eye morphologies in humans appear to be adapted to enhance the particular signals such as gaze direction (Kobayashi & Kohshima, 2001). Thus, the active viewing for eyes by humans may be the behavioural adaptation to enhance their unique forms of facial communications by means of eyes. Likewise, the characteristic patterns of face scanning in chimpanzees may also have been adapted for the communicative purposes, especially for the purpose of the quick, efficient retrieval of emotional information from faces. However, we doubt this latter possibility because their characteristic patterns of eye movements may not derive from the active viewing for particular facial features, as mentioned earlier. Furthermore, the salient (large) mouth morphologies in chimpanzees appear to be adapted for the other purposes than communicative signals, such as food processing (Lambert, 1999).

No matter what adaptive reasons or selective pressures have worked on the characteristic patterns of face scanning in chimpanzees and humans, it should be noted that both patterns seem to have respective advantages in the species-specific forms of facial communications. The persistent and prolonged eye-viewing exhibited by humans might enable the constant retrieval of subtle information conveyed by the eye region. As mentioned earlier, facial muscles and unique eye morphologies contribute to the formation of various subtle expressions in the eye regions of humans. Thus, it would appear to be advantageous for humans to remain focused on the eye region to recognize those species-specific forms of expressions in the eye region. Indeed, it is known that autistic children, who experience difficulties in forming normal social and emotional relationships with people, pay less attention to the eye regions than do children with typical developmental courses (Dalton et al., 2005; Dawson et al., 2004; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Pelphrey et al., 2002). The quick scanning of faces by chimpanzees, on the other hand, might enable them to coarsely but quickly retrieve overall information (e.g., identity, emotion) from faces. As mentioned earlier, compared to humans, face-to-face communications among chimpanzees tend to

occur in brief spurts and during arousing situations. In addition, the mouth regions constitute the main sources of emotional expressions in chimpanzees. Thus, the characteristic patterns of face scanning in chimpanzees ―the quick, vertical scanning for faces, the frequent inspection of the mouth, and the responsiveness to facial actions― would appear to be advantageous to efficiently retrieve general as well as emotional information from faces. The rapid inspection of faces might also be advantageous for chimpanzees with regard to appeasing the other individuals in that it is more likely that prolonged eye contact functions as a threat signal in non human primates than in humans (Gomez, 1996; Thomsen, 1974). These findings suggest an intimate connection between face scanning and species-specific forms of facial communications in chimpanzees and humans, rendering their respective eye movement strategies functional in interactions with conspecifics.

The direct causes for these characteristic scanning patterns in chimpanzees and humans remain unclear; further studies are necessary. For example, the reluctance of eye-viewing by chimpanzees might be due to the less powerful incentive for chimpanzees, compared with humans, to collect eye information, but might also be due to the more powerful incentive for chimpanzees to avoid eye contact. The active viewing for eyes by humans might have been developed to enhance the species-specific forms of emotional communications in which the eye regions play unique roles, as have discussed above. It is also possible that such tendency in humans had been developed for the non-emotional, instrumental communications, in which facial expressions including the eyes have language-like functions. The eye movements signal not only emotions but also the direction of attention or interests. The importance of joint attentional and communicative interactions in humans may have shaped the unique morphologies of eyes (Tomasello et al., 2007) and also the behavioural tendency to actively collect eye information. From the ontogenetical perspective, the active viewing for eyes by humans might derive from biologically determined patterns reflecting certain selective pressures in evolution, but might also constitute patterns learned during the course of development.

Studies in human infants have shown that infants dramatically increased fixations on the eye region at about seven weeks of age (Haith, Bergman, & Moore, 1977). It is known that the differential patterns involving direct/averted gaze emerge early in life (within a few days/weeks) in humans (Farroni, Csibra, Simion, & Johnson, 2002), chimpanzees (Myowa-Yamakoshi, Tomonaga, Tanaka, & Matsuzawa, 2003), and monkeys (Mendelson et al., 1982). However, precise measurements of eye movements have not yet been conducted in infant chimpanzees, and direct comparisons involving these primate infants are necessary to clarify the evolution and development of eye contact.

Because the experiments of this study were conducted in the absence of communicative contexts, these results are most likely to reflect the general species differences which are relatively consistent across various communicative situations. It remains unknown how those characteristic patterns of face scanning in chimpanzees and humans respond to the various communicative contexts; further studies are necessary. Unlike monkey species, chimpanzees are known to often engage in relatively long bouts of eye contacts in affiliative interactions as humans do (Goodall, 1986). They alternate their eye gaze between the food and the experimenter to obtain the foods, perhaps with communicative intent (Leavens & Hopkins, 1998).

The absence of communicative contexts in these experiments might have promoted the general responses to the faces rather than the specific responses to the particular faces (e.g., familiar/unfamiliar species, phylogenetically close/distant species) and also promoted the species differences rather than the group differences within the species. Although the inspection of individual data in this study suggested several behavioural trends of each individual across experiments including previous ones (Kano & Tomonaga, 2009), it did not suggest the specific group differences among them (e.g. social rank, age, sex). Likewise, the previous studies in humans have revealed that the East Asians tended to look at the eye regions for shorter durations than did West Caucasians when presented with photographs of East Asian and West Caucasian faces (Blais,

Jack, Scheepers, Fiset, & Caldara, 2008). Given that human participants of this study (all Japanese) showed the stronger attention to the eye regions than did the chimpanzees, such cultural difference appears to be smaller than the species difference between chimpanzees and humans.

Several similarities between humans and chimpanzees, such as the intense scanning of main facial features and the order in which each facial feature was inspected, are also consistent with the studies on monkeys (Ghazanfar, Nielsen, & Logothetis, 2006; Gothard et al., 2004; Guo et al., 2003). Of special interest in this regard is the characteristic pattern of human face scanning involving sequential re-fixations on the eye region. This pattern is consistent with results of previous human studies (Althoff & Cohen, 1999; Henderson et al., 2005; Walker-Smith et al., 1977). Humans and monkeys are known to look at the eye region longer than at the mouth region, as mentioned earlier. However, this was not the case for the chimpanzees in this study, probably due to the frequent fixations on the mouth region by the chimpanzees. Interestingly, preliminary comparisons with previously published data on monkeys suggested that humans viewed the eye region for the longest durations among these three primate species, that chimpanzees viewed the eye region for as long as did monkeys, and that chimpanzees viewed the mouth region for the longest durations (compare Table 2-4-1 with, for example, Guo et al., 2003). Perhaps the important aspects of human face scanning include the prolonged eye-viewing (or the sequential re fixations on the eye regions) and not simply the dominance of the eyes over the mouth as measured in total viewing time. Additional studies are necessary to clarify this issue.

References

Adachi, I., Chou, D. P., & Hampton, R. R. (2009). Thatcher effect in monkeys demonstrates conservation of face perception across primates. *Curr. Biol., 19*(15), 1270-1273.

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *J Exp Psychol Learn Mem Cogn, 25*(4), 997-1010.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press

Bard, K. A., Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., Costall, A., & Matsuzawa, T. (2005). Group differences in the mutual gaze of chimpanzees (*Pan troglodytes*). *Dev Psychol, 41*(4), 616-624.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3*(8), e3022.

Carpenter, M., & Tomasello, M. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Soc Dev, 4*(3), 217-237.

Chevalier-Skolnikoff, S. (1973). Facial expression of emotion in nonhuman primates. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review* (pp. 11-89). New York: Academic Press.

Dahl, C. D., Wallraven, C., Bulthoff, H. H., & Logothetis, N. K. (2009). Humans and macaques employ similar face-processing strategies. *Curr Biol, 19*(6), 509-513.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nat Neurosci, 8*, 519-526.

Darwin, C. (1999). *The expression of the emotions in man and animals (P. Ekman, Ed)*. London: Fontana. (Original work published 1872).

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Dev Psychol, 40*(2), 271-282.

Ekman, P., & Friesen, W. V. (1978). *The facial action coding system (FACS): A technique for the measurement of facial action*. Palo Alto: Consulting Psychologists Press.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proc Nat Acad Sci, 99*(14), 9602.

Ghazanfar, A. A., Nielsen, K., & Logothetis, N. K. (2006). Eye movements of monkey observers viewing vocalizing conspecifics. *Cognition, 101*(3), 515-529.

Gomez, J. C. (1996). Ostensive behavior in great apes: The role of eye contact. In A. E. Russon, K. A. Bard & S. T. Parker (Eds.), *Reaching into thought: The minds of the great apes* (pp. 131-151). New York: Cambridge University Press.

Goodall, J. (1968). The behaviour of free-living chimpanzees in the Gombe stream reserve. *Anim Behav Monog, 1*(3), 161-311.

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge: Belknap Press.

Gothard, K. M., Erickson, C. A., & Amaral, D. G. (2004). How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Anim Cogn, 7*(1), 25-36.

Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y., & Young, M. P. (2003). How do monkeys view faces?—a study of eye movements. *Exp Brain Res, 150*(3), 363-374.

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science, 198*(4319), 853-855.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Mem Cogn, 33*(1), 98-106.

Itakura, S., & Tanaka, M. (1998). Use of experimenter-given cues during object-choice tasks by chimpanzees (*Pan troglodytes*), an orangutan (*Pongo pygmaeus*), and human infants (*Homo sapiens*). *J Comp Psychol, 112*(2), 119-126.

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: a comparative eye-tracking study. *Proc Roy Soc B, 276*(1664), 1949-1955.

Keating, C. F., & Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces. *Perception, 11*(2), 211-219.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch Gen Psychiat, 59*(9), 809-816.

Kobayashi, H., & Kohshima, S. (1997). Unique morphology of the human eye. *Nature, 387*(6635), 767-768.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J Hum Evol, 40*(5), 419-435.

Lambert, J. E. (1999). Seed handling in chimpanzees (*Pan troglodytes*) and redtail monkeys (*Cercopithecus ascanius*): implications for understanding hominoid and cercopithecine fruit-processing strategies and seed dispersal. *Am J Phys Anthropol, 109*(3), 365-386.

Leavens, D. A., & Hopkins, W. D. (1998). Intentional communication by chimpanzees: A cross-sectional study of the use of referential gestures. *Dev Psychol, 34*(5), 813.

Martin-Malivel, J., & Okada, K. (2007). Human and chimpanzee face recognition in chimpanzees (*Pan troglodytes*): Role of exposure and impact on categorical perception. *Behav Neurosci, 121*(6), 1145-1155.

Matsuzawa, T., Tomonaga, M., & Tanaka, M. (2006). *Cognitive development in chimpanzees.* Tokyo: Springer.

Mendelson, M. J., Haith, M. M., & Goldmanrakic, P. S. (1982). Face scanning and responsiveness to social cues in infant rhesus monkeys. *Dev Psychol, 18*(2), 222-228.

Michel, C., Rossion, B., Han, J., Chung, C. S., & Caldara, R. (2006). Holistic processing is finely tuned for faces of one's own race. *Psychol Sci, 17*(7), 608-615.

Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M., & Matsuzawa, T. (2003). Preference for human direct gaze in infant chimpanzees (*Pan troglodytes*). *Cognition, 89*(2), 113-124.

Nahm, F. K. D., Perret, A., Amaral, D. G., & Albright, T. D. (1997). How do monkeys look at faces? *J Cogn Neurosci, 9*(5), 611-623.

Okamoto, S., Tomonaga, M., Ishii, K., Kawai, N., Tanaka, M., & Matsuzawa, T. (2002). An infant chimpanzee (*Pan troglodytes*) follows human gaze. *Anim Cogn, 5*(2), 107-114.

Parr, L. A., Dove, T., & Hopkins, W. D. (1998). Why faces may be special: evidence of the inversion effect in chimpanzees. *J Cogn Neurosci, 10*(5), 615-622.

Parr, L. A., & Heintz, M. (2008). Discrimination of faces and houses by rhesus monkeys: the role of stimulus expertise and rotation angle. *Anim Cogn, 11*(3), 467-474.

Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying chimpanzee facial expressions using muscle action. *Emotion, 7*(1), 172-181.

Pascalis, O., & Bachevalier, J. (1998). Face recognition in primates: a cross-species study. *Behav Process, 43*(1), 87-96.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *J Autism Dev Disord, 32*(4), 249-261.

Sato, N., & Nakamura, K. (2001). Detection of directed gaze in rhesus monkeys (*Macaca mulatta*). *J Comp Physiol, 115*(2), 115-121.

Thomsen, C. E. (1974). Eye contact by non-human primates toward a human observer. *Anim Behav, 22*, 144-149.

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *J Hum Evol, 52*(3), 314-320.

Tomonaga, M. (1994). How laboratory-raised Japanese monkeys (*Macaca fuscata*) perceive rotated photographs of monkeys: Evidence for an inversion effect in face perception. *Primates, 35*(2), 155-165.

Tomonaga, M. (1999). Inversion effect in perception of human faces in a chimpanzee (*Pan troglodytes*). *Primates, 40*(3), 417-438.

Tomonaga, M. (2007a). Is chimpanzee (*Pan troglodytes*) spatial attention reflexively triggered by the gaze cue? *J Comp Psychol, 121*(2), 156-170.

Tomonaga, M. (2007b). Visual search for orientation of faces by a chimpanzee (*Pan troglodytes*): face-specific upright superiority and the role of facial configural properties. *Primates, 48*(1), 1-12.

van Hooff, J. A. R. A. M. (1967). The facial displays of the catanhine monkeys and apes. In D. Morris (Ed.), *Primate Ethology* (pp. 7-68). Piscataway, NJ: Aldine Transaction.

van Hooff, J. A. R. A. M. (1973). A structual analysis of the social behaviour of a semi-captive group of chimpanzees. In M. von Cranach & I. Vine (Eds.), *Social communication and movement* (pp. 75-162). London: Academic Press.

van Lawick-Goodall, J. (1967). Mother-offspring relationships in free-ranging chimpanzees. In D. Morris (Ed.), *Primate Ethology* (pp. 287-346). Piscataway, NJ: Aldine Transaction.

Vick, S. J., Waller, B. M., Parr, L. A., Pasqualini, M. C. S., & Bard, K. A. (2007). A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (FACS). *J Nonverb Behav, 31*(1), 1-20.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception, 6*(3), 313.

Yarbus, A. L. (1967). Eye movements during perception of complex objects *Eye Movements and Vision* (pp. 171-211). New York: Plenum Press.

Tables

Table.2-4-1 Cumulative viewing time and average fixation duration for the face region and facial features

	Chimpanzee faces			Human faces			Mammal faces		
	Chimpanzees	Humans	*t*-test (*p*)	Chimpanzees	Humans	*t*-test (*p*)	Chimpanzees	Humans	*t*-test (*p*)
Cumulative viewing time (ms)									
Eyes	404.69	1051.73	< 0.001	244.51	884.15	< 0.001	247.43	869.73	< 0.001
Nose	183.45	234.69	0.52	259.00	409.80	0.25	-	-	-
Mouth	265.11	137.56	0.02	161.47	82.81	0.05	220.57	131.53	0.05
Whole face	1037.37	1508.48	< 0.001	900.02	1450.29	< 0.001	1005.57	1501.83	< 0.001
Viewing time as percentage of face viewing time (%)									
Eyes	38.81	69.77	< 0.001	28.07	60.89	< 0.001	23.95	57.96	< 0.001
Nose	18.20	15.50	0.61	28.03	28.13	0.99	-	-	-
Mouth	24.87	9.07	< 0.001	17.80	5.79	< 0.001	21.99	8.73	< 0.001
Average fixation duration (ms)									
Eyes	254.70	456.04	< 0.001	221.40	459.42	< 0.001	217.97	534.59	< 0.001
Nose	246.70	306.61	0.17	275.99	408.21	0.04	-	-	-
Mouth	213.23	277.82	0.05	228.99	294.08	0.15	251.69	319.63	0.02
Whole face	232.63	418.59	< 0.001	251.96	451.58	< 0.001	219.86	426.08	< 0.001

Note. df=22

Figure captions

Figure 2-4-1. (a) Examples of face stimuli presented. The defined regions of interests are indicated by blue lines. (b) Examples of eye movements by chimpanzees and (c) by humans. Each circle represents a fixation that is linked to the adjacent fixation by a line. A longer fixation is drawn as a larger circle. Faces were presented for 2 s. Also, see Movie 1 for a demonstration of these eye movements.

Figure 2-4-2. The similarities and differences in face-scanning patterns between chimpanzees and humans. (a) The number of fixations (mean + SEM) on each feature. (b) The probability of fixation as a function of fixation order.

Figure 2-4-3. The number of fixations as a proportion of the total numbers of face fixations (mean + SEM) on each feature of the facial expressions for chimpanzees and humans. The data were normalized to correct for the differences between facial expressions with regard to the proportions of features.

Figures

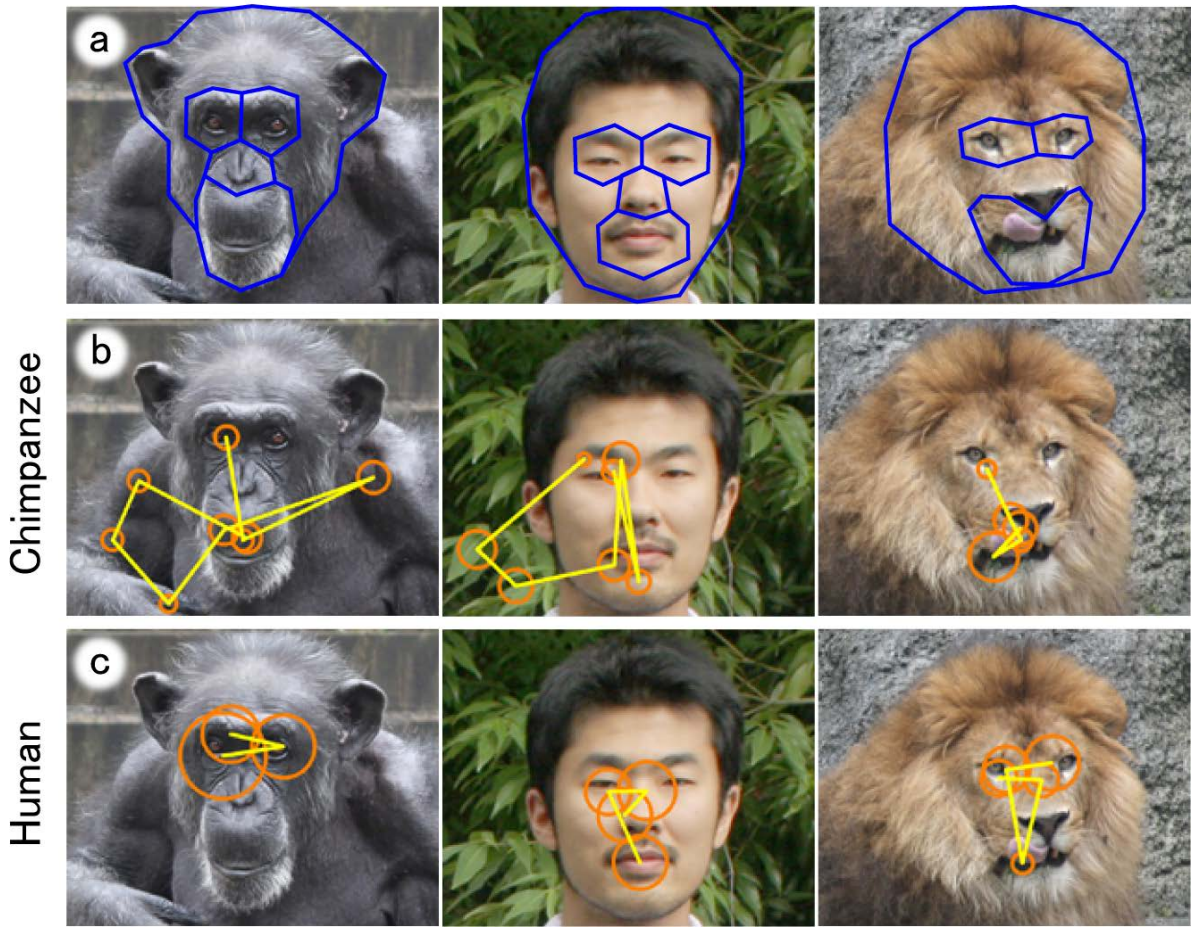


Figure 2-4-1

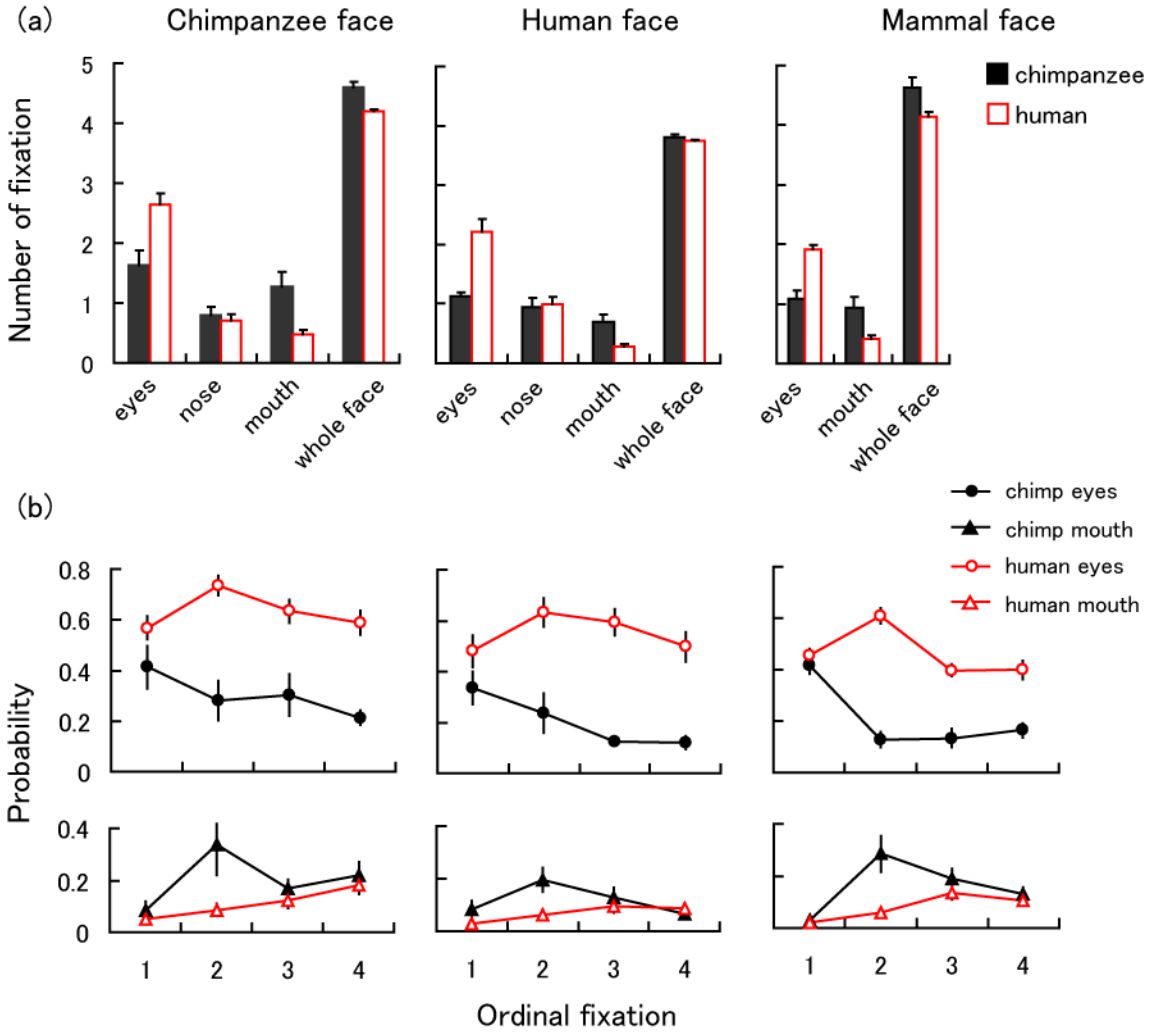

(a)
Chimpanzee face
Human face
Mammal face
chimpanzee
human
Number of fixation
0
1
2
3
4
5
eyes
nose
mouth
whole face
(b)
chimp eyes
chimp mouth
human eyes
human mouth
Probability
0
0.2
0.4
0.6
0.8
1
2
3
4
Ordinal fixation

Figure 2-4-2

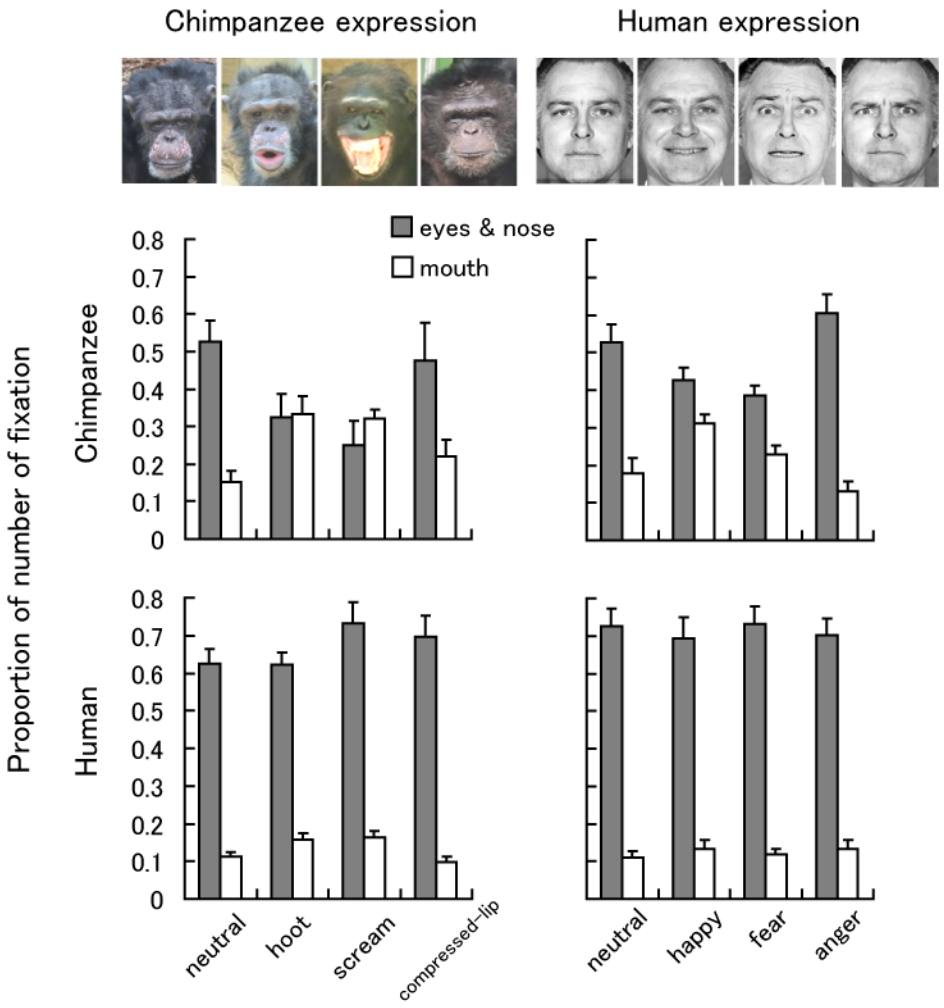

Chimpanzee expression
Human expression
eyes & nose
mouth
Proportion of number of fixation
Chimpanzee
Human
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
neutral
hoot
scream
compressed-lip
neutral
happy
fear
anger

Figure 2-4-3

Chapter 2-5.

The visual strategy specific to humans among hominids: a study using the gap-overlap paradigm

Introduction

Eye-gaze movement constitutes one of the most comprehensively studied visually guided behaviors displayed by humans and macaque monkeys. The visual strategy common to human and nonhuman primates involves the alternation of fixation and saccade; fixation involves maintaining certain parts of the visual field fixed on the fovea, which optimizes retinal acuity and color sensitivity, whereas saccades involve bringing new parts of the visual field onto the fovea using rapid eye movements. Given that primates retrieve visual information primarily from the fovea, how they move their gaze inform us about the ways in which visual information from the external world is retrieved and processed, an operation that is critically important to survival.

When primates shift their gaze from one location to another via saccadic eye movements, competition occurs between two mutually exclusive activities: fixation and saccade initiation. Resolving this competition consumes time because it involves various perceptual and cognitive processes (Findlay & Walker, 1999). This time-consuming competitive process can be examined using a simplified experimental arrangement known as the gap–overlap paradigm. An extensive body of literature exists with regard to the use of this paradigm in humans (Braun & Breitmeyer, 1988; Kalesnykas & Hallett, 1987; Saslow, 1967) and macaque monkeys (Baizer & Bender, 1989; Fischer & Boch, 1983; Fischer & Weber, 1993). Following this paradigm, a central (fixated) and a peripheral target stimulus appear sequentially on a computer screen under two conditions. The central fixation stimulus disappears after a short period of time (200–400 ms) before the target is presented under the gap condition, whereas the central fixation stimulus remains under the overlap condition. The time between target presentation and initiation of a saccade directed at the target is then measured (i.e., the saccade latency). In humans and monkeys, the saccade latency in response to peripheral stimuli has tended to be longer under the overlap than under the gap condition (known as the “gap effect”).

One well-established model of saccade generation (Findlay & Walker, 1999) assumes that resolution of the competition between fixation and saccadic activities requires the integration of various competing information signals to decide whether and where a saccade should occur. This model suggests that resolving this competition involves a relatively slow buildup in one activity and a decline in the other. Thus, when the saccadic activity overcomes the fixation activity, a saccade is generated. The reduction of fixation activity is termed disengagement. Physiologically, this competitive interaction can be observed in a subcortical area, the *superior colliculus,* where a decline in fixation neurons and a buildup in saccade-related neurons occur (Dorris & Munoz, 1995; Dorris, Pare, & Munoz, 1997; Munoz & Wurtz, 1993a, 1993b). Saccade generation is also controlled by various cortical areas including the parietal and frontal cortex (Müri et al., 1998; Munoz & Everling, 2004), especially the frontal eye field (Dias & Bruce, 1994; Hanes & Schall, 1996). According to Findlay and Walker's (1999) model, the gap effect occurs because the fixation activity is automatically reduced by the offset of the fixation stimulus under the gap but not the overlap condition. The offset of the fixation stimulus under the gap condition also works as a warning signal that provides temporal information about the appearance of the target (L. E. Ross & Ross, 1980; S. M. Ross & Ross, 1981).

It is well known that human infants from 1 to 4 months of age have difficulty in shifting their gaze to peripherally presented stimuli, the so called "obligatory fixation" (Stechler & Latz, 1966). The gap–overlap paradigm has revealed that infants in their first year of life show an earlier maturation of saccade latency under the gap than under the overlap condition (i.e., a larger gap effect in younger infants) (Farroni, Simion, Umilt, & Barba, 1999; Hood & Atkinson, 1993; M. Matsuzawa & Shimojo, 1997). Thus, it is suggested that human infants have difficulty in disengaging attention or reducing fixation activity under the overlap condition, in which such disengagement does not occur in an automatic manner. Physiologically, this phenomenon can be explained by the earlier

maturation of subcortical compared with cortical regulatory systems (e.g., the frontal eye field; (Johnson, 1990). Similar difficulties with disengagement under the gap–overlap paradigm have also been reported among individuals with autism (Landry & Bryson, 2004)

Numerous studies have been conducted in macaque monkeys to examine the neural and behavioral mechanisms underlying saccade generation. Lesion studies, behavioral testing, functional neuroimaging studies, single-unit recordings, and anatomical studies in macaques and humans have shown that the neural circuitry controlling saccadic eye movements is homologous, or qualitatively similar, in the two species (Munoz & Everling, 2004). Studies comparing human and nonhuman primates have reported that trained macaques demonstrated a shorter saccadic latency than did trained humans (Baizer & Bender, 1989). During free viewing of a naturalistic dynamic scene, macaques scanned the scene more rapidly than did humans by shifting their gaze to the next location at an earlier time (Berg, Boehnke, Marino, Munoz, & Itti, 2009; Shepherd, Steckenfinger, Hasson, & Ghazanfar, 2010).

A similar species difference was observed in comparisons of humans with one of their closest living primate relatives, chimpanzees (*Pan troglodytes*), when freely viewing static scenes (Kano & Tomonaga, 2009). A subsequent study (Kano & Tomonaga, 2011b) confirmed that this species difference in the timing of gaze movements did not depend on the nature of the stimuli (a scene containing humans/chimpanzees, fruit trees, only background, or texture) and thus seemingly reflected general patterns of gaze movements rather than specific responses to particular components of scenes. That subsequent study also examined the pattern of gaze movements in chimpanzees and humans using the gap–overlap paradigm under free-viewing conditions (no instruction/training) and found that chimpanzees and humans showed very similar saccadic latencies under the gap condition, but that chimpanzees shifted their gaze to the peripheral target at an earlier time than did humans under the overlap condition (i.e., a smaller gap effect in chimpanzees). The species similarity under

the gap condition suggested that perceptual and motor abilities for making a saccade were comparable in both species, and the species differences under the overlap condition suggested the operation of differential visual strategies for resolving the competition between fixation and initiation of a saccade. In this context, it might be argued that humans follow a different pattern than do other primates in the timing of their gaze movements and that this species difference may derive from humans' specific visual strategy for dealing with the aforementioned competition.

Despite their value, the current data have several shortcomings. First, there is the issue of the representativeness of the existing samples. Kano and Tomonaga (2011b) compared six chimpanzees with 18 humans. However, the inclusion of additional individuals would be necessary to confirm that these results reflect species rather than individual differences. This is particularly important considering that these particular chimpanzees were previously extensively trained in computerized tasks, some of which required rapid responses to stimuli presented on a screen. Although these subjects were never trained to make saccades, confirmation of the validity of the free-viewing paradigm as a way to reveal spontaneous viewing patterns would require replication of the aforementioned results with chimpanzees with different training experiences.

Second, there is the issue of the type of stimuli presented in the tasks. Using the gap–overlap paradigm, Kano and Tomonaga (2011b) presented naturalistic figures, faces, and objects rather than simple geometric figures to attract the apes' and humans' spontaneous attention to the stimuli. They found a minimal effect of different types of stimuli on species differences in saccade latencies, even though both species discriminated faces from objects in their gaze responses (in an experimental situation facilitating competition between the two stimulus types). Thus, it was suggested that species differences reflected general (or habitual) patterns of saccade generation rather than the immediate outcomes of the processing of meaningful stimuli. However, one might

argue that faces and objects are both meaningful and that another type of stimulus, such as a meaningless figure, would be necessary to confirm these findings.

Finally, the investigation of species of great apes other than chimpanzees can contribute to clarifying the evolution of gaze-scanning patterns. As indicated above, the ways in which primates move their gaze can inform us about how they retrieve and process visual information that may be critically important for their survival. Thus, it is expected that each species' gaze-scanning pattern sensitively reflects the phylogenetical and socioecological constraints specific to each species. Phylogenetically, chimpanzees are the closest to humans, followed by gorillas and orangutans. Thus, the comparison with the other apes may clarify whether the differences between chimpanzees and humans represent derived or ancestral traits. Additionally, each species has a differential socioecological background (Clutton-Brock & Harvey, 1977). Thus, our comparative study may also help us to assess the potential impact of socio-ecological variables on the gaze-scanning patterns.

The aim of the current study was to examine the timing of gaze movements from comparative perspective using the gap–overlap paradigm. We tested humans and three non-human great ape species, chimpanzees, gorillas and orangutans, living in three different facilities. Additionally, we investigated the effect of the type of stimulus. Following previous studies, we presented faces and objects, but also included a meaningless figure (texture).

Method

Participants

Four female gorillas (one adult, one infant, and two juveniles) and seven orangutans (one adult male, one infant male, four adult females, one juvenile female) housed at the Wolfgang Köhler Primate Research Center (WKPRC) at the Leipzig Zoo in Germany, eight chimpanzees (two adult males, three adult females, three juvenile females) housed at the Great Ape Research Institute (GARI) at Hayashibara Biomedical Laboratories, Inc. in Japan, and eight chimpanzees (two adult males, six adult females) housed at the Primate Research Institute at Kyoto University (KUPRI) participated in this study. Additionally, 16, six, and 18 humans (all adults) were recruited from WKPRC (all Europeans; six males, 10 females), GARI (all Japanese; two males, four females), and KUPRI (all Japanese; six males, 12 females), respectively, to participate in this study. Thus, 27 apes and 40 humans participated in this study. The data from six of the eight chimpanzees and the 18 humans at KUPRI were previously published (Kano and Tomonaga, 2011b). Two additional chimpanzees (an adult male and an adult female) were tested at KUPRI to increase the number of participants. All apes lived in social groups in a large outdoor compound attached to an indoor residence with regular feedings, enrichment, and water *ad libitum*. All apes were neither food- nor water-deprived. All apes and humans voluntarily participated in the study. Animal husbandry and research at WKPRC complied with the "European Association of Zoos and Aquaria (EAZA) Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria" and the "World Association of Zoos and Aquariums (WAZA) Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums," respectively. Animal husbandry at GARI and KUPRI complied with the "Care and Use of Laboratory Animals of Hayashibara Biochemical Laboratories, Inc." and the 2002 version of the "Guidelines for the Care and Use of Laboratory Primates of the Primate Research Institute, Kyoto University," respectively. Research conducted at GARI and

KUPRI was approved by the Animal Welfare and Care Committee of KUPRI and the Animal Research Committee of Kyoto University. Informed consent was obtained from all human participants.

Chimpanzees housed at KUPRI had extensive experience with participation in computerized tasks using a touch-panel display, typically 15–21 inches in size, that required them to respond by touching geometrical or naturalistic figures appearing on the screen (T. Matsuzawa, Tomonaga, & Tanaka, 2006). Chimpanzees at GARI also had experience participating in such touch-panel experiments, but to a lesser extent than those at KUPRI (Idani & Hirata, 2006). Gorillas housed at WKPRC had begun participation in touch-panel experiments only recently, and orangutans at this facility had neither experienced such experiments nor been exposed to images on a computer screen. None of the apes or humans had been explicitly trained to shift their gaze rapidly.

Apparatus

The same eye-tracking techniques were used for apes and humans to ensure the same eye-tracking accuracy among species (Figure 2-5-1 a–c). However, we slightly modified the experimental arrangement in each facility to compensate for the specific constraints and capitalize on the particular resources already present at each institution. Eye-tracking experiments had been previously established with the chimpanzees at GARI and KUPRI (Hattori, Kano, & Tomonaga, 2010; Hirata, Fuwa, Sugama, Kusunoki, & Fujita, 2010; Kano & Tomonaga, 2009, 2010, 2011a, 2011b) and had recently been introduced to apes at WKPRC (Kano, Call, & Tomonaga, in prep.). All apes were tested in an experimental booth. The eye-tracking apparatus and experimenter were located outside the booth and were separated from the apes via transparent acrylic panels at WKPRC and KUPRI. At GARI, the apparatus and experimenter, who was highly familiar to the apes, stayed inside the booth. An eye tracker with an infrared corneal reflection system measured participants' gaze movements. We used a table-mounted apparatus at WKPRC and KUPRI (60 Hz; Tobii X120,

Tobii Technology AB, Stockholm, Sweden) and a monitor-integrated type at GARI (60 Hz; Tobii T60), both of which were based on the same technology and thus possessed the same eye-tracking performance. These eye trackers were equipped with wide-angle lenses (±40 degrees in the semicircle above the eye camera) and recorded both eyes, thereby allowing relatively large head movements by participants. The eye tracker and 17-inch LCD monitor (1280 × 1024 pixels at WKPRC and KUPRI and 1024 × 768 pixels at GARI) were mounted on a movable platform, and the distance between the platform and the participants was adjusted to the point at which gaze was most accurately recorded (approximately 60 cm). This flexible adjustment of the distance between the platform and the participants enabled us to record the gaze movements of apes without any head-restraining device. Four of the 11 apes (all juveniles) at WKPRC had difficulty approaching the panel upon the request of experimenter. For this reason, we used a nozzle and a tube attached to the panel, which continuously produced drops of grape juice during the experiment, thereby keeping the participants' heads in front of the panel. Although they were sipping grape juice during the presentation of stimuli, they did not attend to the nozzle but freely moved their eyes. The experimenter at GARI sat beside the apes and held their heads lightly during the recordings. The other apes at WKPRC and KUPRI sat still in front of acrylic panels, and the experimenter encouraged them to face the eye tracker. The apes received small pieces of fruit *ad libitum* before and after the calibration procedure and presentation of pictures. No reward was given to reinforce any particular gaze behavior. At KUPRI, humans were tested in the same experimental booth as apes, whereas humans were tested in another room at GARI and WKPRC. Although the eye tracker recorded the eyes of humans and apes at KUPRI and those of the apes at WKPRC through the transparent acrylic panel, we confirmed that the acrylic panels (1.5–2 cm thick, absent of dirt or scratches) had no influence on the eye-tracking data in the preliminary test for accuracy. Each participant's gaze was recorded as a relative coordinate with respect to the monitor size (i.e., not as

the gaze angle). One degree of gaze angle corresponded to approximately 1 cm on the screen at a typical 60-cm viewing distance.

Figure 2-5-1

An automated sequential calibration procedure was conducted for both apes and humans. Five-point calibration was used for humans, but the calibration points were reduced to two for apes to avoid interruption of the automated calibration process by participants averting their gaze elsewhere. We asked humans to fixate on the small dot appearing on each calibration point. For apes, we presented a small object or image at the calibration point for apes (a piece of fruit or a small video clip; approx. 1-2 degree in width/height), thereby drawing their spontaneous attention to the point. The calibration was repeated for the apes until maximum accuracy was obtained. The calibration accuracy was checked *post-hoc* by presenting a small object or image at several points on the screen and manually monitoring the participants' gaze toward those points. To reduce the time required for a daily session, the same calibration data were used for apes on separate days when the same level accuracy was achieved at the beginning of a daily session. To avoid any calibration error due to changes in posture or eye surface, the calibration accuracy was checked several times during the daily session, and the calibration was repeated when the same accuracy was not obtained. To quantitatively estimate the positional error, we conducted a preliminary session for each ape and human, in which we recorded the position of the participant's gaze on the small object or image. We then calculated the distance between the center of object/image and the recorded gaze position. The error was found to be within 0.5–0.7-degree, on average, for all groups; this was sufficiently accurate for the requirements of this study. Daily sessions lasted for 10–15 minutes for each ape and human.

Procedure

Each trial began after participants focused on a small red mark appearing at a central position on the screen. We then presented a central fixation stimulus followed by a target stimulus (approx. 4.8 × 4.8°at a typical 60-cm viewing distance, approx. 9°apart). The target appeared randomly to the left or right 560 ms after the onset of the trial (Figure 2-5-1d). We measured the time between target presentation and the initiation of a saccade directed at the target (i.e., the saccade latency). Under the gap condition, the central fixation stimulus disappeared 260 ms before target presentation, whereas the central fixation stimulus remained under the overlap condition. The peripheral target stimulus remained for 940 ms, and thus each trial lasted 1.5 s in total. Two types of stimulus, faces and objects, were initially used to test the GARI and KUPRI groups. Another stimulus type, texture, was also used to test the WKPRC group. Face stimuli included both ape and human faces. We prepared more than 50 exemplars of each stimulus type. Different exemplars of the same stimulus type were presented at both central and peripheral locations within each trial. Each exemplar was randomly selected from the entire pool of exemplars. A previous study conducted at KUPRI (Kano and Tomonaga, 2011b) involved six trials under each condition for each stimulus type (6×2×2 = 24 trials in total). Because that study confirmed the minimal variance across trials, we reduced the number of trials to three under each condition for each stimulus type at WKPRC (3×2×3 = 18 trials in total) and GARI (3×2×2 = 12 trials in total).

We randomized the presentation order of conditions and stimulus type for each participant. The entire session was conducted on a single day for humans at GARI (12 trials) and KUPRI (24 trials) and on two separate days for humans at WKPRC (nine trials each day). Six trials were conducted each day for apes at all facilities (3, 2, and 4 days in total at WKPRC, GARI, and KUPRI, respectively). Preliminary analysis, however, revealed no significant effect of day among those apes and humans tested on separate days.

After the completion of the whole session, we repeated trials in which participants prematurely shifted their gaze before the onset of the peripheral target. If the same occurred in those repeated trials, we excluded those trials from the analysis. This procedure resulted in the total data loss of 0.0%, 1.3% and 19.8% of all trials for humans, gorillas, and orangutans, respectively, at WKPRC; 6.9% and 11.4% for humans and chimpanzees, respectively, at GARI; and 0.2% and 2.0% for humans and chimpanzees, respectively, at KUPRI. We found no bias for a particular stimulus type or condition in those excluded trials. Additionally, for the quantitative analysis, we excluded the trials in which the saccade latency of participants was longer than the average for all trials (281 ms) plus 2.5 standard deviations (274 ms; i.e., longer than 555 ms) or in which the participants did not shift their gaze by the end of a trial. This resulted in the total data loss of 7.6%, 1.4%, and 0.0% of all trials for humans, gorillas and orangutans, respectively, at WKPRC; 4.4 and 4.7% for humans and chimpanzees, respectively, at GARI; and 4.8% and 1.5% for humans and chimpanzees, respectively, at KUPRI. These trials appeared primarily under the overlap condition, which probably reflects a characteristic of that condition, as discussed below. We found no bias for a particular stimulus type in these excluded trials.

Results

Figure 2-5-2 presents the saccade latency of participants as a function of stimulus type (face, object, and texture) and condition (gap and overlap). We conducted three separate analyses of variance (ANOVAs) for the WKPRC, GARI, and KUPRI groups. We found a significant interaction between condition and species in the WKPRC group (gorillas, orangutans, and humans) ($F_{2, 24}$ = 11.10, $P < 0.001$, $\eta^2 = 0.48$), which was explained by the difference in saccade latency among the species being more evident under the overlap ($F_{2, 24} = 23.15$, $P < 0.001$, $\eta^2 = 0.65$) than under the gap ($F_{2, 24} = 3.92$, $P = 0.033$, $\eta^2 = 0.24$) condition. Comparisons between orangutans and humans, between gorillas and humans, and between gorillas and orangutans showed interactions between condition and species that were significant ($F_{1, 21} = 19.50$, $P < 0.001$, $\eta^2 = 0.48$), not significant ($F_{1, 18}$ = 2.12, $P = 0.16$, $\eta^2 = 0.10$), and marginally significant ($F_{1, 9} = 4.80$, $P = 0.056$, $\eta^2 = 0.34$), respectively. The analysis by species showed that the effect of condition was significant for humans ($F_{1, 15} = 159.97$, $P < 0.001$, $\eta^2 = 0.91$) and gorillas ($F_{1, 3}$= 615.80, $P < 0.001$, $\eta^2 = 0.99$) and marginally significant for orangutans ($F_{1, 6} = 4.54$, $P = 0.07$, $\eta^2 = 0.43$). The effect of stimulus type was not significant, either main effect or interaction with condition ($P > 0.05$).

Figure 2-5-2

We found a significant interaction between condition and species at GARI (chimpanzees and humans) ($F_{1, 12} = 14.51$, $P = 0.002$, $\eta^2 = 0.54$), which can be explained by the species difference in saccade latency being more evident under the overlap ($F_{1, 12} = 7.72$, $P = 0.017$, $\eta^2 = 0.39$) than the gap ($F_{1, 12} = 3.77$, $P = 0.076$, $\eta^2 = 0.23$) condition. The analyses by species showed that the effect of condition was significant for humans ($F_{1, 5} = 24.17$, $P = 0.004$, $\eta^2 = 0.82$) but not for chimpanzees ($F_{1, 7} = 0.71$, $P = 0.42$, $\eta^2 = 0.093$). The effect of stimulus type was not significant, either main effect

or interaction with condition ($P > 0.05$).

We found a significant interaction between condition and species at KUPRI (chimpanzees and humans) ($F_{1, 24} = 15.72$, $P = 0.001$, $\eta^2 = 0.39$), which can be explained by the species difference in saccade latency being more evident under the overlap ($F_{1, 24} = 8.38$, $P = 0.008$, $\eta^2 = 0.25$) than under the gap ($F_{1, 24} = 0.002$, $P = 0.96$, $\eta^2 < 0.001$) condition. The analyses by species showed that the effect of condition was significant for both humans ($F_{1, 17} = 59.05$, $P = 0.004$, $\eta^2 = 0.77$) and chimpanzees ($F_{1, 7} = 12.25$, $P = 0.010$, $\eta^2 = 0.63$), although the difference was small for chimpanzees (13 ms). We found a significant effect of stimulus type ($F_{1, 24} = 5.69$, $P = 0.025$, $\eta^2 = 0.19$) in the KUPRI group, although the difference was small (8 ms).

We did not have a sufficient number of individual samples to examine the effects of age and sex, but the exclusion of juveniles and males did not alter the aforementioned statistical results. The effect of laterality (either right or left) was not significant in terms of either main effects or interactions ($P > 0.05$) at any of the facilities. We confirmed the same statistical results even when we limited the analyses to the first six trials at all facilities.

Figure 2-5-3 presents the distribution of saccade latencies from 0 to 500 ms. Apes and humans showed a skewed distribution in their saccade latencies, with the peaks around 200 ms. The distributions were similar across species under the gap condition. In contrast, the distribution was skewed more leftward in apes than in humans, and the distribution of the saccade latencies in humans was characterized by a long right tail exceeding 300 ms under the overlap condition. This species difference was most pronounced between humans and chimpanzees/orangutans, and the results for gorillas were between these two extremes.

Figure 2-5-3

Table 2-5-1

Discussion

Humans, chimpanzees, gorillas, and orangutans from three research facilities were compared in terms of their saccade latencies using the gap–overlap paradigm. Although all species showed similar saccade latencies under the gap condition, the species clearly differed from one another under the overlap condition. In general, humans showed longer saccade latencies did than the other apes under that condition, which is explained by saccade latencies longer than 300 ms being more frequent in humans than in other apes. We found little evidence that stimulus type affected latencies.

The similarities among the various species under the gap condition suggest common perceptual and motor abilities for responding to peripherally presented stimuli (e.g., the sensitivity to salience of peripheral vision, the execution of saccadic eye movements). Relatively consistent saccade latencies under the gap condition have also been observed among human infants at various developmental stages (Hood & Atkinson, 1993; M. Matsuzawa & Shimojo, 1997) and in autistic and typically developing children (Landry & Bryson, 2004). These phylogenetic, developmental, and clinical consistencies suggest a relatively primitive or fundamental operation underlying saccade generation under the gap condition.

In contrast, humans and apes differed under the overlap condition. Unlike chimpanzees and orangutans, humans showed a clear overlap (or gap) effect. Interestingly, gorillas showed a pattern that was somewhat similar to that displayed by humans in this regard. Given that the species were similar in their saccade latencies under the gap condition, the species differences under the overlap condition suggest the use of differential strategies for dealing with the competition between fixation and initiation of a saccade. An alternative account is that the participants anticipated the appearance of the target *via* the offset of the fixation stimulus (or the gap) (Reuter-Lorenz, Hughes, & Fendrich, 1991; L. E. Ross & Ross, 1980). However, this account is unlikely in the context of this

study because any of our participants were not trained in the task, as will be discussed in greater detail below.

Our study investigated participants' spontaneous (or "natural") pattern of gaze shifting rather than their ability to control their gaze. Thus, they viewed the stimuli freely without any instruction or training and without any head-restraining device. This arrangement differed critically from those used in previous studies with instructed/trained humans and macaques as subjects (e.g., (Fischer & Boch, 1983; Fischer & Ramsperger, 1984) and resembles those used in previous studies with untrained human infants as subjects (Hood & Atkinson, 1993; M. Matsuzawa & Shimojo, 1997). Several lines of evidence indicate that the participants in this study showed their spontaneous patterns of gaze shifting. First, differing amounts of experience in participating in computerized tasks or exposure to computer screens did not affect the data obtained from apes. Second, the different reward schedule for apes (receiving a reward for participating in the experiments, but not for their gaze behaviors) did not affect the results of this study. Third, we observed few express or anticipatory saccades (fewer than 100 ms), phenomena that have been frequently observed in trained subjects when the location at which the target appeared was predictable (Fischer & Weber, 1993). Finally, analysis of the first six trials of the session, in which an effect for (uninstructed) training or learning was unlikely, yielded results identical to those for the entire session. Interestingly, the untrained humans in this study showed a skewed distribution of saccade latency, with a long right tail extending beyond 300 ms under the overlap condition, which has been commonly observed in humans who were freely viewing naturalistic scenes. In contrast, trained humans in the previous studies have often shown a symmetrical or inverted bell-shaped distribution of saccade latency under the overlap condition (Braun & Breitmeyer, 1988; Fischer & Weber, 1993; Reuter-Lorenz, et al., 1991).

Despite its theoretical importance, the free viewing design of this experiment may also

have shortcomings given that the possible differences in the motivational states of each species may have had certain influence on the results (although we did not find any behavioral evidence to show such motivational differences). We also recognize that genuine natural patterns of gaze movements can be observed only during the course of daily activities. Thus, further studies simulating naturalistic contexts (e.g., use of head-mounted eye-tracking devices) (Land, Mennie, & Rusted, 1999) are necessary.

We found no effect of stimulus type on saccade latency. Thus, the saccade latency elicited by the gap–overlap paradigm in this study may have reflected a habitual or well-automated process for saccade generation rather than an immediate outcome of processing meaningful stimuli. Kano & Tomonaga (2011b) found a similar result in chimpanzees and humans. That is, when an object and (a seemingly more salient) face were presented at central and peripheral locations, respectively, under one condition and in the opposite locations, respectively, under the other condition, the saccade latencies of both species were shorter when the object was presented centrally and the face was presented peripherally than *vice versa*, whereas the effect of overlap remained the same under both conditions. Therefore, although the saccade latency of participants seems to be influenced by the stimulus type, the effect of overlap seems to be influenced by the competition between the two stimuli, rather than by the stimulus type *per se*.

Somewhat surprisingly, among the nonhuman great apes, gorillas showed the clearest overlap effect. However, the small sample size (n = 4) precludes our reaching a definitive conclusion, and further studies are necessary to confirm this result. One interpretation for this possible species difference is the possible behavioral or cognitive uniqueness of gorillas among the great apes, which has been suggested by previous studies (Peignot & Anderson, 1999; Suarez & Gallup, 1981). Alternatively, gorillas may have been somewhat neophobic to the presented stimuli. That is, their attention (or effective visual field) may have been temporarily narrowed to the central stimuli,

rendering them less sensitive to the appearance of peripheral stimuli.

At least two ultimate (or evolutionary) interpretations are possible with regard to the benefits (and costs) of the adoption of such specific visual strategies by humans and apes. First, the specific visual strategy used by each species may have survival value in specific socioecological environments. For example, it may be more beneficial to scan visual fields more quickly by shifting gazes earlier in the context of arboreal living, where objects and animals tend to appear in an unexpected manner, as may be the case for chimpanzees and orangutans. To clarify the effect of socioecological factors, additional comparative studies in various primate species are necessary.

Second, the pronounced effect of overlap (or competition) in humans may reflect their unique means of information processing among hominids. That is, rather than constantly retrieving new information, humans may keep their gaze stationary and thereby promote time-consuming internal processing (e.g., for the sake of categorical and language processing). In contrast, apes may switch their focus of attention (i.e., the fovea) more frequently than humans and may thereby cover a wider visual field via gaze movements. Thus, a trade-off between the depth and breadth of information processing/retrieval may occur in human and nonhuman apes. However, two limitations must be considered with regard to this hypothesis. First, no quantitative information is available about the information retrieval/processing in these species in this study. Second, the hypothesis does not explain the effect of the overlap in gorillas (although the effect was somewhat smaller in gorillas than in humans). One could assume that this overlap effect in gorillas derived from a different cause than that in humans, as explained above; however, further studies are necessary to clarify this issue.

In conclusion, this study found phylogenetic similarities and differences in saccade latencies among hominid species. Although all species seem to have similar perceptual and motor mechanisms for performing saccades, the species may differ in their strategies for coping with the competition between two activities involving fixation and saccade initiation. In particular, humans

seem to spend a longer time resolving this competition than the other great apes, which may reflect this species' unique means of information processing.

References

Baizer, J. S., & Bender, D. B. (1989). Comparison of saccadic eye movements in humans and macaques to single-step and double-step target movements. *Vision Res., 29*(4), 485-495.

Berg, D. J., Boehnke, S. E., Marino, R. A., Munoz, D. P., & Itti, L. (2009). Free viewing of dynamic stimuli by humans and monkeys. *J. Vision, 9*(5), 1-15.

Braun, D., & Breitmeyer, B. G. (1988). Relationship between directed visual attention and saccadic reaction times. *Exp. Brain. Res., 73*(3), 546-552.

Clutton-Brock, T. H., & Harvey, P. H. (1977). Primate ecology and social organization. *J. Zool., 183*(1), 1-39.

Dias, E. C., & Bruce, C. J. (1994). Physiological correlate of fixation disengagement in the primate's frontal eye field. *J. Neurophysiol., 72*(5), 2532-2537.

Dorris, M. C., & Munoz, D. P. (1995). A neural correlate for the gap effect on saccadic reaction times in monkey. *J. Neurophysiol., 73*(6), 2558-2562.

Dorris, M. C., Pare, M., & Munoz, D. P. (1997). Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. *J. Neurosci., 17*(21), 8566-8579.

Farroni, T., Simion, F., Umilt, C., & Barba, B. D. (1999). The gap effect in newborns. *Dev. Sci., 2*(2), 174-186.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behav. Brain Sci., 22*(04), 661-674.

Fischer, B., & Boch, R. (1983). Saccadic eye movements after extremely short reaction times in the monkey. *Brain Res., 260*(1), 21-26.

Fischer, B., & Ramsperger, E. (1984). Human express saccades: extremely short reaction times of goal directed eye movements. *Exp. Brain. Res., 57*(1), 191-195.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behav. Brain Sci., 16*(03),

553-567.

Hanes, D. P., & Schall, J. D. (1996). Neural control of voluntary movement initiation. *Science, 274*(5286), 427-430.

Hattori, Y., Kano, F., & Tomonaga, M. (2010). Differential sensitivity to conspecific and allospecific cues in chimpanzees and humans: A comparative eye-tracking study. *Biol. Lett., 6*(5), 610-613.

Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., & Fujita, S. (2010). Facial perception of conspecifics: chimpanzees (*Pan troglodytes*) preferentially attend to proper orientation and open eyes. *Anim. Cogn., 13*(5), 679-688.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Inf. Behav. Dev., 16*(4), 405-422.

Idani, G., & Hirata, S. (2006). Studies at the Great Ape Research Institute, Hayashibara. . In D. A. Washburn (Ed.), *Primate perspectives on behavior and cognition* (pp. 29-36). Washington: American Psychological Association.

Johnson, M. H. (1990). Cortical maturation and the development of visual attention in early infancy. *J. Cogn. Neurosci., 2*(2), 81-95.

Kalesnykas, R. P., & Hallett, P. E. (1987). The differentiation of visually guided and anticipatory saccades in gap and overlap paradigms. *Exp. Brain. Res., 68*(1), 115-121.

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: a comparative eye-tracking study. *Proc. Roy. Soc. B., 276*(1664), 1949-1955.

Kano, F., & Tomonaga, M. (2010). Face scanning in chimpanzees and humans: continuity and discontinuity. *Anim. Behav., 79*(1), 227-235.

Kano, F., & Tomonaga, M. (2011a). Perceptual mechanism underlying gaze guidance in chimpanzees and humans. *Anim. Cogn., 14*(3), 377-386.

Kano, F., & Tomonaga, M. (2011b). Species difference in the timing of gaze movement between chimpanzees and humans. *Anim. Cogn., 14*(6), 879-892.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311-1328.

Landry, R., & Bryson, S. E. (2004). Impaired disengagement of attention in young children with autism. *J. Child Psychol. Psychiat., 45*(6), 1115-1122.

Müri, R. M., Rivaud, S., Gaymard, B., Ploner, C. J., Vermersch, A. I., Hess, C. W., et al. (1998). Role of the prefrontal cortex in the control of express saccades. A transcranial magnetic stimulation study. *Neuropsychologia, 37*(2), 199-206.

Matsuzawa, M., & Shimojo, S. (1997). Infants' fast saccades in the gap paradigm and development of visual attention. *Inf. Behav. Dev., 20*(4), 449-455.

Matsuzawa, T., Tomonaga, M., & Tanaka, M. (2006). *Cognitive development in chimpanzees*. Tokyo: Springer.

Munoz, D. P., & Everling, S. (2004). Look away: the anti-saccade task and the voluntary control of eye movement. *Nature Rev. Neurosci., 5*(3), 218-228.

Munoz, D. P., & Wurtz, R. H. (1993a). Fixation cells in monkey superior colliculus. I. Characteristics of cell discharge. *J. Neurophysiol., 70*(2), 559-575.

Munoz, D. P., & Wurtz, R. H. (1993b). Fixation cells in monkey superior colliculus. II. Reversible activation and deactivation. *J. Neurophysiol., 70*(2), 576-589.

Peignot, P., & Anderson, J. R. (1999). Use of experimenter-given manual and facial cues by gorillas (*Gorilla gorilla*) in an object-choice task. *J. Comp. Psychol., 113*(3), 253-260.

Reuter-Lorenz, P. A., Hughes, H. C., & Fendrich, R. (1991). The reduction of saccadic latency by prior offset of the fixation point: an analysis of the gap effect. *Atten. Percept. Psychophys., 49*(2), 167-175.

Ross, L. E., & Ross, S. M. (1980). Saccade latency and warning signals: Stimulus onset, offset, and change as warning events. *Atten. Percept. Psychophys., 27*(3), 251-257.

Ross, S. M., & Ross, L. E. (1981). Saccade latency and warning signals: effects of auditory and visual stimulus onset and offset. *Atten. Percept. Psychophys., 29*(5), 429-437.

Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *J. Opt. Soc. Am., 57*(8), 1024-1029.

Shepherd, S. V., Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human-monkey gaze correlations reveal convergent and divergent patterns of movie viewing. *Curr. Biol., 20*(7), 649-656.

Stechler, G., & Latz, E. (1966). Some observations on attention and arousal in the human infant. *J. Am. Acad. Child Psychiatry, 5*, 517-525.

Suarez, S. D., & Gallup, G. G. (1981). Self-recognition in chimpanzees and orangutans, but not gorillas. *J. Hum. Evol., 10*(2), 175-188.

Tables

Table 2-5-1

Distribution of Saccade Latency (ms) of Individuals. Each individual engaged in six, nine, and 12 trials respectively at WKPRC, GARI, and KUPRI.

	151–200	201–250	251–300	301–350	351–400	401-450	total
Gap condition							
WKPRC							
Human		7	9				16
Gorilla	1	2	1				4
Orangutan	2	5					7
GARI							
Human		4	2				6
Chimpanzee		2	4	2			8
KUPRI							
Human	3	12	3				18
Chimpanzee	1	5	1	1			8
Overlap condition							
WKPRC							
Human				5	8	3	16
Gorilla			2	1	1		4
Orangutan		4	2	1			7
GARI							
Human			2	2		2	6
Chimpanzee		1	5	2			8
KUPRI							
Human		4	4	6	2	2	18
Chimpanzee		6	1	1			8

Figure captions

Figure 2-5-1. (a)–(c) an ape on an eye tracker at WKPRC (Wolfgang Köhler Primate Research Center, Germany), GARI (Great Ape Research Institute, Japan), and KUPRI (Primate Research Institute, Kyoto University, Japan), respectively. (d) Gap–overlap paradigm used in this study.

Figure 2-5-2. Saccade latency (ms) as a function of condition (gap and overlap) and stimulus type (face, object, and texture) in humans, chimpanzees, gorillas, and orangutans from three research facilities. Error bars represent ±95% confidence intervals.

Figure 2-5-3. Frequency distribution of saccade latencies from 0 to 500 ms in four species from three research facilities. The data were pooled for all participants and stimulus types. The bin was 50 ms.

Figures

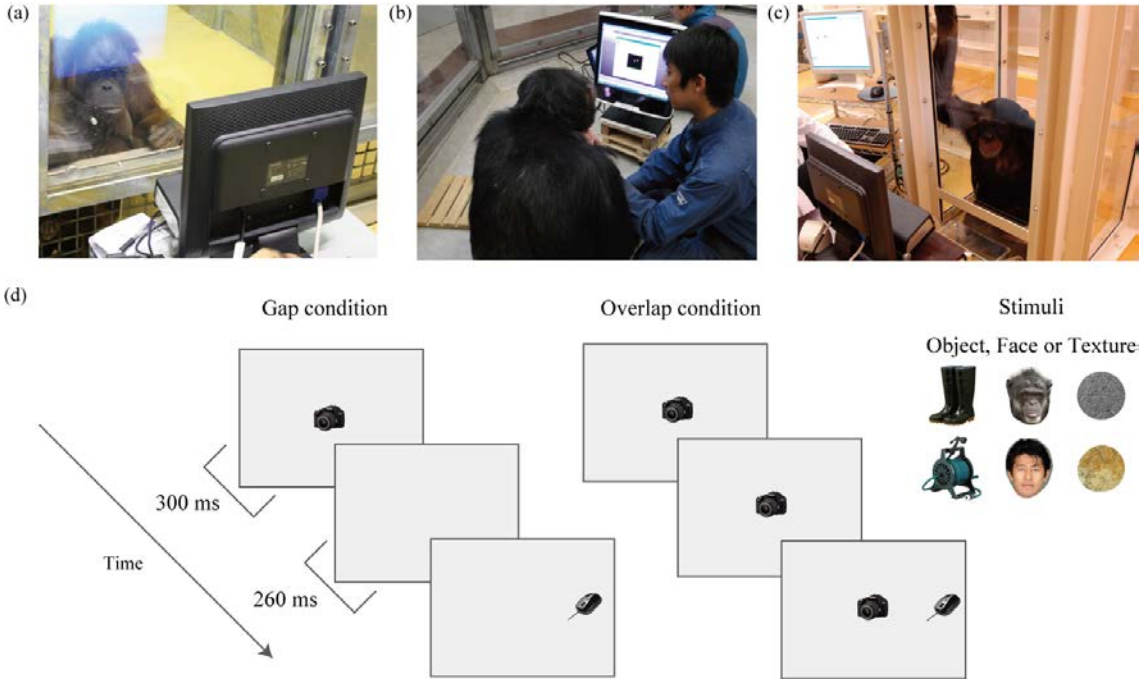


Figure 2-5-1

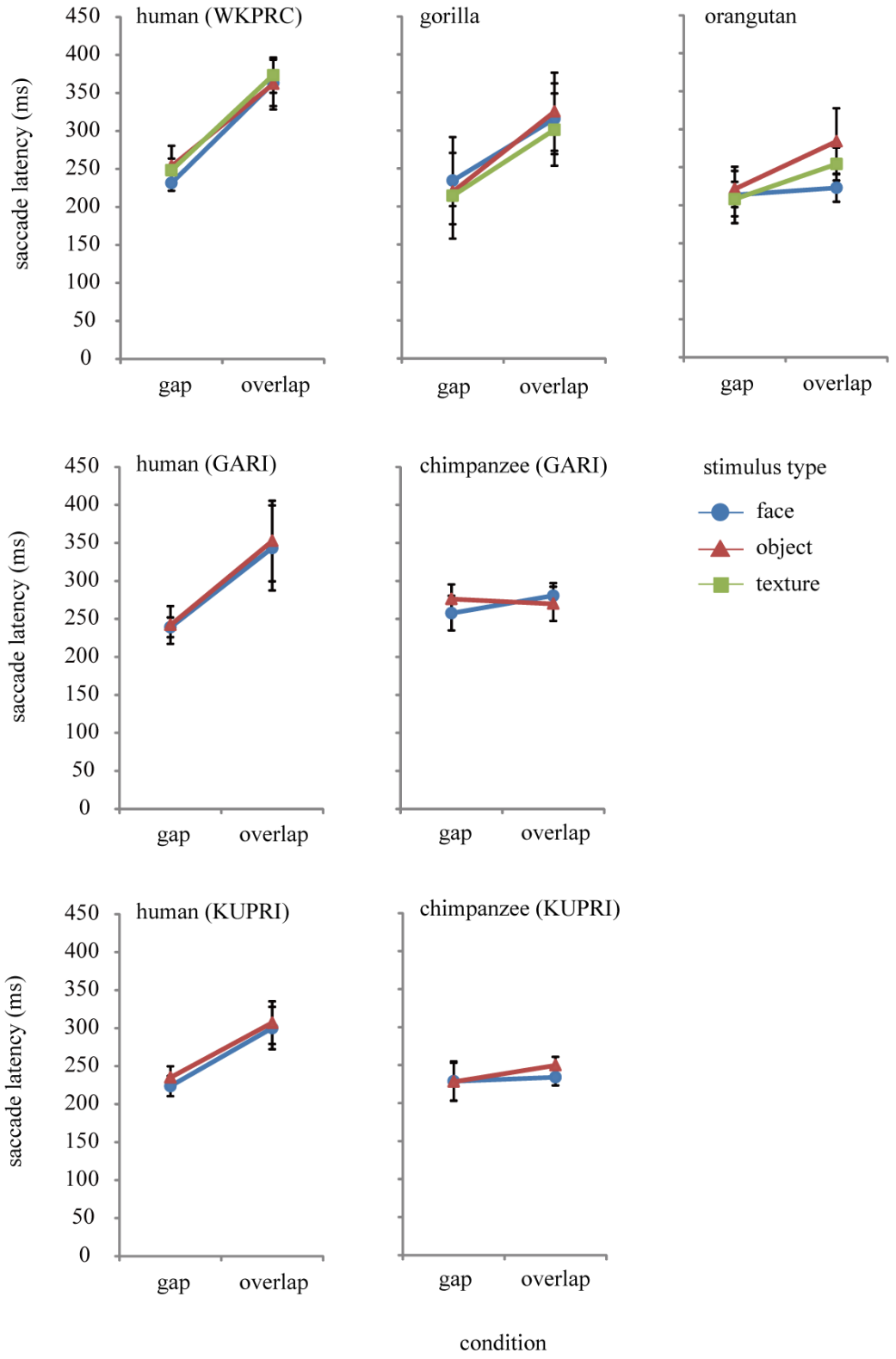

human (WKPRC)
gorilla
orangutan
human (GARI)
chimpanzee (GARI)
human (KUPRI)
chimpanzee (KUPRI)
saccade latency (ms)
450
400
350
300
250
200
150
100
50
0
gap
overlap
condition
stimulus type
face
object
texture

Figure 2-5-2

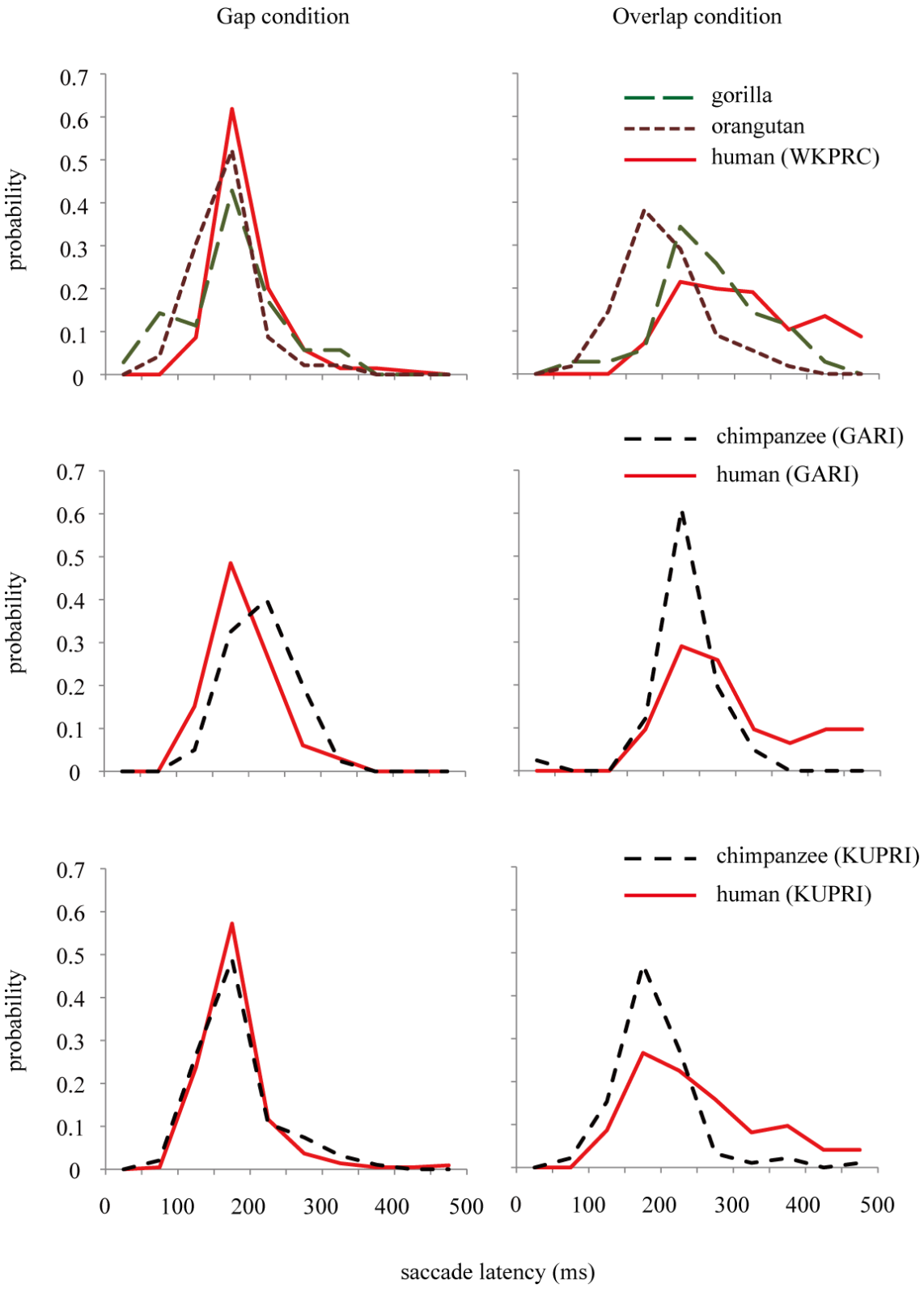

Gap condition
Overlap condition
gorilla
orangutan
human (WKPRC)
chimpanzee (GARI)
human (GARI)
chimpanzee (KUPRI)
human (KUPRI)
probability
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
100
200
300
400
500
saccade latency (ms)

Figure 2-5-3

Chapter 2-6.

Face scanning in four genera of hominid: an eye-tracking study in humans, chimpanzees, gorillas, and orangutans

Introduction

Faces convey a rich variety of information about social partners, including age, sex, identity, emotions, and intentions. Measuring gaze movements during face scanning is a useful way to examine how individuals retrieve and process the information contained in faces. Studies of human adults have revealed that the pattern of face scanning is dependent on the type and difficulty of the task (Armann & Bülthoff, 2009; Heisz & Shore, 2008; Malcolm, Lanyon, Fugard, & Barton, 2008). In an extreme case, participants instructed not to move their eyes during face learning exhibited a clear deficit in later recognition of those faces (Henderson, Williams, & Falk, 2005).

Developmental studies have found that human newborns prefer to track face-like stimuli over non-face stimuli with their heads and eyes (Johnson, Dziurawiec, Ellis, & Morton, 1991; Valenza, Simion, Cassia, & Umilt, 1996); similar findings have been reported for monkeys and chimpanzees (Kuwahata, Adachi, Fujita, Tomonaga, & Matsuzawa, 2004; Myowa-Yamakoshi, Yamaguchi, Tomonaga, Tanaka, & Matsuzawa, 2005). From 5 to 7 weeks of age, human infants exhibit an increased tendency to fixate on eyes and a decreased tendency to fixate on the periphery of faces (Hainline, 1978; Haith, Bergman, & Moore, 1977). Clinical studies have found that autistic participants view the inner feature of faces (i.e., eyes, nose, and mouth), especially eyes, less frequently than control participants do (Dalton et al., 2005; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Pelphrey et al., 2002 but see van der Geest, Kemner, Verbaten, & van Engeland, 2002). Somewhat similar atypical patterns of scanning internal features, especially eyes, have also been reported among schizophrenic and socially phobic participants (Delerue, Lapr vote, Verfaillie, & Boucart, 2010; Green, Williams, & Davidson, 2003; Horley, Williams, Gonsalvez, & Gordon, 2004; Phillips & David, 1997).

Comparative studies have shown that, like humans, macaques look predominantly at eyes, more than at the nose and mouth (Guo, Robertson, Mahmoodi, Tadmor, & Young, 2003; Keating &

Keating, 1982; Mendelson, Haith, & Goldmanrakic, 1982; Nahm, Perret, Amaral, & Albright, 1997). Macaques also alter their patterns of face scanning in a similar manner to humans when the features and configurations of presented faces are altered, suggesting that macaques have at least some similarities to humans in their face-processing strategies (Adachi, Chou, & Hampton, 2009; Dahl, Wallraven, Bulthoff, & Logothetis, 2009; Gothard, Brooks, & Peterson, 2009; Guo, et al., 2003; Keating & Keating, 1993)); similar results have also been reported for chimpanzees (Hirata, Fuwa, Sugama, Kusunoki, & Fujita, 2010).

Direct, quantitative comparisons between chimpanzees and humans have revealed specific similarities and differences in face-scanning patterns between these closely related species (Kano and Tomonaga, 2009, 2010). The two species are similar in that they view faces for longer period and at an earlier time (at the first and second fixation) than other parts of the body. Additionally, they both view the inner features of faces, especially eyes, for longer period and at an earlier time than the periphery of faces. However, the two species differ quantitatively in their viewing patterns of faces and eyes. Humans are more likely than chimpanzees to maintain their gaze on the face and eyes once they have fixated. Chimpanzees are more likely than humans to shift their gaze to other parts of the body or face. The similarities between chimpanzees and humans highlight the homologous nature of face scanning between the two species, and the differences between chimpanzees and humans reveal the eye-viewing pattern specific to humans.

Currently, face-scanning data are not available for gorillas and orangutans, the other two great apes. Therefore, it is unclear whether the similarities and differences between humans and chimpanzees are indicative of general similarities and difference between humans and nonhuman great apes or whether isolated similarities and differences exist among the hominid species. Gorillas and orangutans are phylogenetically close to both chimpanzees and humans, but they are relatively more distant from humans than chimpanzees are. Additionally, although all these species are highly

skilled at social interaction, the natural grouping patterns of gorillas and orangutans are socially less specialized than are those of chimpanzees and humans. In particular, western lowland gorillas form stable groups of one male and multiple females, whereas orangutans lead semi-solitary lives, with mother–offspring pairs forming the only long-lasting associations. Chimpanzees show a fission–fusion grouping pattern, forming both large groups, including multiple males and multiple females, and smaller stable subgroups. Based on these differences, gorillas and orangutans may be expected to exhibit patterns of face scanning that differ from human patterns more than the patterns exhibited by chimpanzees do. Another possibility is that all great apes share generally similar face-scanning patterns, reflecting their phylogenetic closeness and similarities in social skills, with several exceptional patterns in humans (e.g., eye-viewing pattern).This study examined this issue by investigating face-scanning patterns among gorillas, orangutans, and humans and comparing them to those observed previously in chimpanzees and humans.

We used the eye-tracking method previously developed to assessing the pattern of face scanning in chimpanzees under a free-viewing condition for static faces (Kano & Tomonaga, 2009, 2010). In particular, we assessed which facial feature was scanned most intensely and in what order and how systematically various facial features were scanned. We were also interested in the extent to which each species was sensitive to variations in faces, including individual familiarity and species familiarity (i.e., conspecific vs. allospecific); i.e. the extent to which the typical face-scanning pattern of each species was generalized to a variety of faces. We also presented gorillas and orangutans with sex- and age-specific faces. Gorillas and, especially, orangutans exhibit marked variations in their facial morphology according to sex and age. Adult gorilla males have a high crest at the top of their heads, and adult orangutan males have a developed flange on the sides, both of which are thought to be sexual signals (Ankel-Simons, 2000; Kuze, Malim, & Kohshima, 2005). Infant and juvenile orangutans have pale coloring around their eyes and mouth, thought to signal

immaturity (Kuze, et al., 2005). We examined the extent to which those conspecific facial cues were visually appealing to gorillas and orangutans as measured by their scanning patterns.

Method

Participants

A total of 5 gorillas (*Gorilla gorilla*), 10 orangutans (*Pongo abelii*), and 12 humans (5 males and 7 females; all European adults) participated in this study. All apes were housed in semi-natural indoor and outdoor enclosures (total 2564 m^2 for gorillas, 1910 m^2 for orangutans) at the Wolfgang Köhler Primate Research Center (WKPRC), Leipzig Zoo, Germany; all received regular food, enrichment, and water *ad libitum*. The apes were not deprived of food or water. All apes and humans voluntarily participated in the study. Animal husbandry and research complied with the EAZA Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria and the WAZA Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums, respectively. Informed consent was obtained from all human participants.

We compared our finding with previously-published eye-tracking data from 6 chimpanzees (*Pan troglodytes*) and 18 humans (7 males and 11 females; all Japanese adults), which were collected using similar procedures at the Primate Research Institute, Kyoto University (KUPRI), Inuyama, Japan (Kano & Tomonaga, 2009, 2010). The details of housing conditions, ethics statements, and experimental procedures of the KUPRI apes have been published elsewhere (Kano and Tomonaga, 2009, 2010). Table 2-6-1 lists the sex, age, and rearing history of each WKPRC ape. All apes and humans had extensive experience interacting with both conspecifics and humans and were thus highly familiar with both kinds of faces. The KUPRI chimpanzees, like the human participants, had extensive experience observing photographs of faces (Matsuzawa, Tomonaga, & Tanaka, 2006), but the WKPRC apes had no experience in that regard.

Table 2-6-1

Apparatus

All humans and apes (both WKPRC and KUPRI) used the same apparatus (except for the experimental booth used, see below) to ensure the same accuracy of eye tracking between species. All apes were tested in a chamber that separated the ape from the eye-tracking apparatus and the experimenter with transparent acrylic panels. A table-mounted eye tracker measured eye movement using an infrared corneal reflection technique (60 Hz; Tobii X120, Tobii Technology AB, Stockholm, Sweden). This eye tracker has wide-angle lenses (±40 degrees in the semicircle above the camera) and records both eyes, thereby allowing for relatively large head movements by participants. The eye tracker and the 17-inch LCD monitor (1280 × 1024 pixels) were mounted on a movable platform, and the distance between the platform and the participants was adjusted by the experimenter to the point at which the gaze could be most accurately recorded (approx. 60 cm). This flexible adjustment of the distance between the platform and the participants enabled us to record the gaze movements of chimpanzees without any head restraint. In WKPRC, however, 8 of the 15 apes tested (Gorgo, Louna, Zola, Batak, Kila, Maia, Suaq, and Tanah; mostly juveniles) had difficulty in approaching the acrylic panel upon the request of experimenter. We thus used a nozzle and tube attached to the acrylic panel, which produced regular drops of grape juice during the experiment, thereby keeping their heads in front of acrylic panel (Figure 2-6-1a and b). The other apes sat still in front of the acrylic panel, and the experimenter encouraged them to face the eye tracker. Those apes received small pieces of fruit *ad libitum* before and after the calibration procedure and presentation of pictures (Figure 2-6-1c). The eye tracker was unable to track both eyes of one male adult gorilla (Gorgo) due to the wide distance between his eyes, and of a juvenile female gorilla (Louna) due to her strabismus. Therefore, we concealed the left eye of these apes from the eye tracker by attaching opaque tape to the panel, and tracked only the right eye. Although these apes viewed stimuli only with their right eyes, they still had a full view of all stimuli with the typical 60-cm viewing distance. The preliminary analysis

revealed no specific pattern of face scanning in these two apes, especially relating to laterality, so we did not exclude their data from the analysis. Although ape experiments were conducted for apes in the presence of experimenter, the apes rarely attended to the experimenter during the presentation of stimuli. Also, the apes who received grape juice during the presentation of pictures did not look at the nozzle but freely viewed the pictures. Humans were tested in another room with no panel between the eye tracker and participants. Our preliminary tests for accuracy revealed that the acrylic panels (1.5–2 cm thick, with no scratches or dirt) had no influence on the eye-tracking data. Each participant's gaze was recorded as a relative coordinate with respect to the monitor size (i.e., not as the gaze angle). One degree of gaze angle corresponded to approximately 1 cm on the screen at a typical 60-cm viewing distance.

Figure 2-6-1

An automated five-point calibration was conducted for humans; the calibration points were reduced into two for apes to minimize the time required for each calibration procedure. We asked humans to fixate on a small dot appearing on each calibration point. For apes, we presented a short video clip or a piece of fruit and thereby attracted their attention to the calibration point; this calibration was repeated until maximum accuracy was obtained. To check for accuracy, we presented humans and apes with five small red marks on the screen; we asked humans to fixate on the marks and presented apes with a piece of fruit in front of the mark to attract their attention. For apes, we used the same calibration data for separate sessions (on different days) to reduce the time required for each daily session. To avoid calibration errors due to changes in posture or eye surfaces, an accuracy check was conducted once before and several times during each daily session, and the calibration was repeated when the original accuracy was not maintained. Under these calibration and

accuracy-check procedures, the accuracy was checked for two gorillas, six orangutans, six chimpanzees, and six humans; error (the distance between intended and recorded gaze position) was small and comparable among the various species (average error ± s.e.m. 0.61 ± 0.06, 0.72 ± 0.04, and 0.52 ± 0.05 degrees visual angle at the typical 60-cm viewing distance for gorillas, orangutans, and humans, respectively).

Stimuli

Figures 2-6-2a and b present examples of the stimuli. Color still pictures were prepared for this study, including 16 whole-body and 56 facial pictures. Gorillas and orangutans viewed eight human and eight conspecific whole-body pictures (total of 16 whole-body pictures) and eight human and 24 conspecific facial pictures (32 facial pictures); each ape viewed a total of 48 pictures. Humans viewed pictures of both apes and humans (16 whole-body and 56 facial pictures). Half of the pictures were of familiar apes/humans that interacted with the participants on a daily basis, and the other half included unfamiliar apes/humans that had been never exposed to the participants. Photographs were taken when the subjects were in calm, relaxed states. The facial pictures of gorillas and orangutans consisted of three types of faces (eight stimuli for each type): adult male face, adult female face, and juvenile face. As noted above, adult male gorillas have a high crest on the top of their head, adult male orangutans have a developed flange on the side of their head, and juvenile orangutans have pale coloring around the eyes and mouth. Whole-body pictures were converted to 1280 × 1024 pixels (37 × 30 degrees at a typical 60-cm viewing distance). Facial pictures were converted to 768 × 1024 pixels (22 × 30 degrees at a typical 60-cm viewing distance) with a gray frame around the background (total 1280 × 1024 pixels). The internal parts of faces (eyes, nose, and mouth) were thus approximately 10～15 degrees in size at a typical 60-cm viewing distance.

Figure 2-6-2

Procedure

In each trial, a picture was presented after participants fixated on a red mark that appeared at the center of the screen. Participants then scanned the picture freely (without any training or instruction). They never kept gazing at the point where the initial red mark appeared, and sequential scanning of the picture was almost always observed. Each stimulus was presented for 3 sec. No specific instructions were given to humans except to view the pictures freely.

The presentation order of the pictures was randomized for each participant. The entire session was conducted over 2 days for humans (36 pictures each day), but the session was divided across 12 days for apes (four pictures each day). The purpose of dividing the entire session for apes was to reduce the time required for each daily experiment and to maintain the apes' interest in the pictures. Each daily experiment lasted 10–15 min for apes and 15–20 min for humans. The effect of habituation was assessed after the entire session was completed, but the results were not significant (see Results section). Trials in which participants did not view a picture for longer than 1 sec were repeated after the whole session, and the original trials were replaced by the new trials; otherwise, those trials were eliminated from the analysis. As a result, we excluded 6.2% and 8.7% of trials involving whole-body pictures and 10.0% and 7.1% of trials involving facial pictures for gorillas and orangutans, respectively (no trials were eliminated for humans). The exclusion of trials (especially those involving facial pictures) was largely attributable to two male individuals, Gorgo (gorilla) and Bimbo (orangutan). These individuals sometimes averted their heads from facial pictures, so the results were excluded. The reason for this behavior is uncertain; it could have been active gaze avoidance from the social stimuli or simply a lack of interest in the pictures.

Data analysis

Fixation definition. Fixation was defined as a stationary gaze within a radius of 50 pixels

for at least 75 ms (more than five measurement samples). Otherwise, the recorded sample was defined as part of a saccade. Records during the first 200 ms were eliminated from the analysis, thereby eliminating fixations that might have begun before the onset of stimuli.

Area of interest (AOI). Each stimulus was divided into areas of interest (AOI) for quantitative comparison. Each whole-body picture was divided into background, face, and body. Each AOI was generated to be 20 pixels larger than the precise outline of the features to compensate for error in gaze estimation. The AOIs were placed over each other in the following order: background, body, and face (i.e., face was topmost). If two or more AOIs were duplicated, gaze samples were added to the upper AOI. Each facial picture was divided into background, eyes, nose, mouth, and periphery (ears, cheeks, chin, forehead, hair; see Figure 2-6-2c). In addition to the eye AOI, which included the entire region around the eyes, we defined an eyeball AOI, which included only the eyeballs. Thus, the eyeball AOI is included in the 'around eye' AOI.

Viewing time. The proportion of viewing time for each AOI was calculated with respect to the viewing time for the entire scene, of either whole-body pictures or facial pictures. Any out-of-scene fixations were excluded from analyses of whole-body pictures (less than 5% for all fixations), and any out-of-face fixations were excluded from analyses of facial pictures (1.9, 6.1, and 6.0% of all fixations, respectively, for humans, gorillas, and orangutans). Because the area of each AOI was different, viewing time was normalized for area size by subtracting the proportion of viewing time from the proportion of area size. The chance level was thus set at zero.

Probability of fixation across fixation order. To examine the time course of face viewing in whole-body pictures and eye viewing in facial pictures, we calculated the probability of fixation on faces/eyes across fixation order. The results were calculated as the proportion of fixations with respect to the total number of sampled fixations for each fixation order.

Probability of saccade destination. To examine the typical scanpath for facial pictures, we

calculated the probability of saccade destination. For this analysis, the eye AOI was divided into right and left eye (i.e., right eye, left eye, nose, mouth, and periphery). Each saccade (the path that joins two consecutive fixations) was classified based on the two AOIs in which the saccade started and ended; thus, each saccade was classified into one of 10 possible combinations of beginning and ending AOIs. Saccades for out-of-face start or end points were excluded from analyses. The results were calculated as the proportion of saccades with respect to the total number of sampled saccades.

Edit cost. To assess similarities in typical scanpaths for faces between trials (within each individual) and between individuals (within each species), we calculated the edit costs (Levenstein distance) of two given scanpaths. Edit cost refers to the minimum number of edit procedures (replacement, insertion, deletion, and addition) necessary to match two strings (Foulsham & Underwood, 2008). Scanpaths were defined as the first four consecutive fixations (e.g., eyes–mouth–eyes–nose, nose–eyes–nose–mouth). Saccades for out-of-face start or end points were excluded from analyses. To calculate an edit cost between trials (within individual), we first compared each scanpath with all the other scanpaths of the same individual across trials and then averaged those values for each scanpath. Similarly, to calculate an edit cost between individuals (within species), we first compared each scanpath with the scanpaths of all the other individuals from the same species viewing the same stimulus and averaged those values for each comparison.

Results

First, it is important to note that none of the ape participants exhibited a fear response to any pictures, even though some apes (gorillas and orangutans, especially juveniles) had never been exposed to facial pictures before this study.

Figure 2-6-3

Whole-body pictures

Proportion of viewing time. Figure 2-6-3a shows the proportion of viewing time for each AOI when participants were presented with whole-body pictures. We conducted a repeated-measures ANOVA with species, AOI, stimulus species, and familiarity as factors. Overall, we found a significant main effect of AOI ($F_{2, 48}$ = 446.59, P < 0.001, η^2 = 0.94). The effect of AOI was significant in all species: humans ($F_{2, 22}$ = 382.82, P < 0.001, η^2 = 0.97), gorillas ($F_{2, 8}$ = 50.41, P < 0.001, η^2 = 0.92), and orangutans ($F_{2, 18}$ = 178.01, P < 0.001, η^2 = 0.95). Post hoc tests (Bonferroni correction) revealed that these species viewed faces, bodies, and backgrounds in that order (*Ps* < 0.001). Additionally, we found a significant interaction between species and AOI ($F_{4, 48}$ = 18.64, P < 0.001, η^2 = 0.60); the effect of species was significant in viewing patterns for faces ($F_{2, 24}$ = 28.69, P < 0.001, η^2 = 0.70), bodies ($F_{2, 24}$ = 5.15, P = 0.014, η^2 = 0.30), and backgrounds ($F_{2, 24}$ = 24.50, P < 0.001, η^2 = 0.67). Post hoc tests revealed that humans differed from the two ape species, especially in their strong viewing tendency for faces (*Ps* < 0.001). We could not analyze the effects of age and sex because of the small numbers of juveniles and males. However, excluding juveniles from the analysis did not change the overall pattern of results (main AOI: $F_{2, 40}$ = 396.84, P < 0.001, η^2 = 0.95; species × AOI: $F_{4, 40}$ = 20.20, P < 0.001, η^2 = 0.66). Additionally, excluding males from the analysis did not change the overall patterns (main AOI: $F_{2, 44}$ = 391.20, P < 0.001, η^2 = 0.94; species × AOI:

$F_{4, 44}$ = 17.34, P < 0.001, η^2 = 0.61). To assess the effect of habituation, we separated the whole session into two blocks and included that factor in the ANOVA; however, we did not find any effect of session block (block × AOI: $F_{2, 23}$ = 0.76, P = 0.47, η^2 = 0.06; block × AOI × species: $F_{4, 48}$ = 1.19, P = 0.32, η^2 = 0.09).

These results are consistent with previous findings comparing chimpanzees and humans (Kano & Tomonaga, 2009). Additionally, a comparison between studies (see Table 2-6-2) did not reveal any particular differences in the viewing time for each scene area between the two groups of humans and the three ape species.

Table 2-6-2

We also found a significant interaction between stimulus species and AOI ($F_{2, 48}$ = 4.72, P = 0.013, η^2 = 0.16). The effect of stimulus species was significant in viewing patterns for faces ($F_{1, 24}$ = 10.56, P = 0.003, η^2 = 0.30) and bodies ($F_{1, 24}$ = 6.09, P = 0.021, η^2 = 0.20), which is explained by the fact that all species viewed faces of conspecific pictures longer than those of allospecific pictures. Additionally, we found a significant interaction between familiarity, species, and AOI ($F_{4, 48}$ = 2.93, P = 0.030, η^2 = 0.19). The interaction between familiarity and AOI was significant only in humans ($F_{2, 22}$ = 4.34, P = 0.026, η^2 = 0.28); humans viewed faces of familiar individuals somewhat longer than those of unfamiliar individuals (0.65 vs. 0.60).

Probability of fixation on faces across fixation order. Figure 2-6-3b shows the time course of face-viewing patterns for whole-body pictures. A repeated-measures ANOVA with species and fixation order as factors revealed a main effect of fixation order ($F_{4, 96}$ = 18.23, P < 0.001, η^2 = 0.43), which is explained by the fact that these species were more likely to fixate faces at early rather than later fixation timings. We also found a significant interaction between species and fixation order ($F_{8,}$

$_{96}$ = 4.22, P < 0.001, η^2 = 0.26), which is explained by the fact that humans were more likely to fixate on faces at an early time than were the two ape species.

Facial pictures

Proportion of viewing time. Figure 2-6-4a shows the proportion of viewing time for each AOI when participants were presented with facial pictures. We conducted a repeated-measures ANOVA with species, AOI, stimulus species, and familiarity as factors. Juvenile faces of gorillas and orangutans were not included in this analysis; i.e. the effect of face type (male, female, juvenile) was assessed in a separate analysis. Overall, we found a significant main effect of AOI ($F_{3, 69}$ = 238.75, P < 0.001, η^2 = 0.91). The effect of AOI was significant in all species, humans ($F_{3, 33}$ = 146.57, P < 0.001, η^2 = 0.93), gorillas ($F_{3, 12}$ = 100.94, P < 0.001, η^2 = 0.96), and orangutans ($F_{3, 24}$ = 65.90, P < 0.001, η^2 = 0.89). Post hoc tests revealed that these species viewed eyes, nose/mouth, and periphery in that order (*Ps* < 0.001). Additionally, we found a significant interaction between species and AOI ($F_{3, 69}$ = 3.14, P = 0.009, η^2 = 0.21); the effect of species was significant in viewing patterns for eyes ($F_{2, 23}$ = 4.00, P = 0.032, η^2 = 0.25) and periphery ($F_{2, 23}$ = 6.27, P = 0.007, η^2 = 0.35). Post hoc tests revealed that humans differed from orangutans in viewing patterns of eyes and periphery (*Ps* < 0.001). Gorillas did not differ significantly from the other two species; their viewing patterns were somewhat between those of humans and orangutans. Narrowing the 'around eye' region to the 'eyeball' region revealed obvious differences in eye-viewing patters by species ($F_{2, 23}$ = 8.09, P = 0.002, η^2 = 0.41). Post hoc tests revealed that humans differed from the two ape species in their strong tendency to view the eyeball region (*Ps* < 0.05), indicating that humans viewed eyes more directly than did the two ape species. We were unable to analyze the effects of age and sex because of the small numbers of juveniles and males. However, excluding juveniles from the analysis did not change the overall pattern of results (main AOI: $F_{3, 57}$ = 174.85, P < 0.001, η^2 = 0.90; species × AOI: $F_{6, 57}$ = 3.20, P = 0.009, η^2 = 0.25), and excluding males from the analysis did not change the overall

pattern of results (main AOI: $F_{3, 66} = 218.51$, $P < 0.001$, $\eta^2 = 0.90$; species × AOI: $F_{6, 66} = 3.13$, $P = 0.009$, $\eta^2 = 0.22$). To assess the effect of habituation, we separated the whole session into two blocks and included that factor in the ANOVA; however, we did not find any effect of session block (block × AOI: $F_{3, 22} = 1.87$, $P = 0.16$, $\eta^2 = 0.20$; block × AOI × species: $F_{3, 22} = 1.49$, $P = 0.20$, $\eta^2 = 0.16$).

Figure 2-6-4

The stronger eye-viewing tendency among humans compared with apes is consistent with previously published comparisons of chimpanzees and humans (Kano & Tomonaga, 2010). A comparison between studies (see Table 2-6-3) did not reveal any particular differences among ape species in viewing time for internal features of faces (or periphery of faces). Although we observed a certain amount of variation among ape species of two studies in viewing time for the region around eyes, narrowing the eye AOI (i.e., eyeball AOI) minimized this species difference (see Table 2-6-3). The two groups of humans differed somewhat in their viewing patterns, especially with regard to eyes and mouth. Humans at KUPRI viewed the eyes for longer durations and the mouth for shorter durations than did humans at WKPRC. This was probably because KUPRI humans tended to fixate more internally on the faces (see Figs. 6 and 7), supporting previous findings comparing Asian and Western study populations (Blais, Jack, Scheepers, Fiset, & Caldara, 2008). Narrowing the eye AOI (i.e., eyeball AOI) minimized human group/cultural differences (see Table 2-6-3).

Table 2-6-3

We found a significant interaction among stimulus species, AOI, and species ($F_{6, 69} = 6.62$, $P < 0.001$, $\eta^2 = 0.36$). This is explained by the fact that, whereas gorillas and orangutans viewed the

eyes of conspecific faces for longer durations than those of allospecific faces, humans exhibited an opposite tendency, tending to view the eyes of allospecific faces for longer durations than those of allospecific faces. The effect of familiarity was not significant in this analysis ($P > 0.05$).

Probability of fixation on eyes across fixation order. Figure 2-6-4b shows the time course of eye-viewing patterns in facial pictures (data for juvenile faces were not included). For the 'around eye' region, a repeated-measures ANOVA using species and fixation order as factors revealed a main effect of fixation order ($F_{4, 96} = 35.03$, $P < 0.001$, $\eta^2 = 0.59$), which is explained by the fact that these species were more likely to fixate eyes at early rather than later fixation timings. The interaction between species and fixation order was not significant ($F_{8, 96} = 1.69$, $P = 0.11$, $\eta^2 = 0.12$). For the 'eyeball' region, we found a significant interaction between species and fixation order ($F_{8, 96} = 2.28$, $P = 0.028$, $\eta^2 = 0.16$). This is explained by the fact that, whereas humans exhibited an increasing tendency for viewing eyeball regions, the two ape species exhibited a decreasing tendency for viewing the same region.

The effect of face type. Figure 2-6-5 shows the proportion of viewing time for each AOI as a function of differential face types in gorillas and orangutans. A repeated-measures ANOVA with face type and AOI did not reveal an effect of face type in gorillas ($F_{6, 24} = 1.97$, $P = 0.10$, $\eta^2 = 0.33$), whereas it did in orangutans ($F_{6, 54} = 6.09$, $P < 0.001$, $\eta^2 = 0.40$). Post hoc tests revealed that orangutans viewed the eyes of juvenile faces for longer durations compared with male or female faces and that they viewed the periphery of male faces (i.e., cheek flange) for longer durations compared with that area of female or infant faces ($Ps < 0.05$; Figure 2-6-5). Human participants who viewed those faces did not exhibit any significant effects of face type (gorilla faces: $F_{6, 66} = 0.08$, $P = 0.99$, $\eta^2 = 0.008$; orangutan faces: $F_{6, 66} = 1.53$, $P = 0.18$, $\eta^2 = 0.12$).

Figure 2-6-5

Probability of saccade destination. Figure 2-6-6 shows the typical scanpaths for faces among human and apes (data for juvenile faces were not included). Data include results from both WKPRC and KUPRI groups. Most strikingly, apes and humans differed in their patterns for scanning left and right eyes. WKPRC humans shifted their gaze between the left and right eyes more frequently than did gorillas ($t_{15} = 2.26$, $P = 0.039$) and orangutans ($t_{20} = 4.82$, $P < 0.001$). KUPRI humans shifted their gaze between the left and right eyes more frequently than chimpanzees ($t_{19} = 3.18$, $P = 0.005$). See Figure 2-6-7 for examples.

Figure 2-6-6

Figure 2-6-7

Edit cost. Table 2-6-4 shows the edit costs of two scanpaths for facial pictures (data for juvenile faces were not included). Overall, the scanpaths were more similar between trials and between individuals in humans than in apes. This means that the scanpaths of humans were generally more stereotyped between trials and between individuals than were those of apes.

Table 2-6-4

Discussion

Our results revealed both similarities and differences in how humans and nonhuman primates scanned pictures of whole bodies and faces. On one hand, all species viewed faces for longer durations and at earlier times than they viewed bodies and backgrounds when presented with whole-body pictures. Additionally, all species viewed inner features of faces, especially eyes, for longer durations and at earlier times than they viewed the periphery when presented with facial pictures. On the other hand, humans differed from apes in that they viewed faces for longer durations than apes when presented with whole-body pictures. Furthermore, humans viewed eye regions more actively than apes did, searched for 'eyeball' regions, and alternated their gaze between the left and the right eyes when presented with facial pictures. These face-scanning patterns of apes and humans were highly consistent across stimuli, despite the fact that various appearances of faces were presented. However, apes altered their scanning pattern somewhat accordingly to the type of faces presented. First, apes viewed conspecific faces and eyes longer than allospecific faces and eyes, although humans did not necessarily follow this pattern. Additionally, orangutans differentiated among sex- and age-related differential facial morphologies of conspecifics in their scanning patterns (humans and gorillas did not exhibit significant differences in this regard).

A striking similarity among all ape species highlighted the homologous rather than species-specific nature of face-scanning patterns. No clear differences appeared among chimpanzees, gorillas, and orangutans; all apes exhibited comparable viewing time of faces when presented with whole-body pictures (Table 2-6-2) and of internal features when presented with facial pictures (Table 2-6-3). The viewing time of each internal feature of faces differed somewhat; for example, gorillas viewed the region around eyes somewhat longer than did the other two ape species (Table 2-6-3). However, this does not mean that gorillas viewed the eye region more actively than the other ape species, as they exhibited similar patterns with regard to direct eye fixation ('eyeball' region; see

Figure 2-6-5). Additionally, although each ape species exhibited somewhat different patterns for the saccade-destination network when viewing conspecific faces (Figure 2-6-6), these differences are largely explained by the fact that differential conspicuities in the faces of each ape species attracted gazes differentially to those regions (e.g., protruded mouth in chimpanzees, large nostrils in gorillas, developed cheek in orangutans).

Clear difference appeared between apes and humans in viewing patterns for the eyes. First, humans tended to fixate directly on eyes (‘eyeball’ rather than ‘around eye’), and that tendency gradually increased during later fixations, whereas apes were more likely than humans to fixate on the region around the eyes throughout the time course (Figure 2-6-4b). Second, humans often alternated their gaze between the left and right eye, whereas apes did not show this gaze movement (Figure 2-6-5). By alternating their gaze between the left and right eyes and shifting their gaze occasionally to the mouth, humans exhibited triangular scanpaths over faces. The absence of the upper component of this triangular shape among apes made their scanpaths appear to be more linear than those of humans. Third, the scanpaths of humans were more stereotyped than those of apes (Table 2-6-4), probably due to their tendency to actively view both eyes (i.e., alternation of gaze at the left and right eye).

One possible explanation for the distinctive human pattern of viewing the eye region is that eyes may have evolved additional communicative functions not found in ape species. Compared with other primates, human eyes have a notable black–white contrast between the iris and sclera (Kobayashi & Kohshima, 2001). Moreover, fine motor control of the muscles around the eyes (Ekman & Friesen, 1978) enables humans to communicate a variety of directional and emotional cues. Thus, active viewing of eyes may benefit humans, by efficiently retrieving such communicative information. That is, humans may employ eyes for bidirectional communication between senders and receivers. An alternative interpretation of the active viewing of eye regions

among humans is that eye regions are more important than other facial features during face processing, and thus humans' intense fixation on eyes reflects their superior ability to process faces. However, this interpretation is weakened by previous findings, which indicate that among humans, recognition of identity is related to intense and systematic fixations on the entire complex of inner facial features, not necessarily the eyes (Hsiao & Cottrell, 2008; Malcolm, et al., 2008; Stacey, Walker, & Underwood, 2005). Furthermore, as mentioned in the Introduction, the level of viewing eyes appears to be influenced by the level of development and clinical symptoms such as autism and schizophrenias. One study found that variation in eye viewing among autistic individuals was strongly and positively associated with amygdala activation, suggesting heightened emotional response associated with eye viewing in autistic individuals (Dalton, et al., 2005). These studies further underscore the communicative function of viewing these regions among humans.

We found that orangutans viewed the periphery of adult male faces (flange on the sides) more intensely than those of adult female or juvenile faces. Furthermore, they viewed the eyes of juvenile faces (surrounded by pale coloring) more intensely than those of adult faces. These results suggest that those facial areas specific to age and sex are visually appealing to orangutans, effectively signaling those meanings to receivers. This differential pattern for each type of face may in part derive from the less stereotyped (more flexible) pattern of face scanning in orangutans than in humans (Table 2-6-3). Similar species contrasts were reported about chimpanzees' and humans' scanning of facial expressions (Kano & Tomonaga, 2010). When chimpanzees and humans were shown static facial expressions of chimpanzees and humans, humans focused on the eyes of both neutral and emotional faces, whereas chimpanzees shifted their attention from the eyes to the conspicuous features of emotional faces (e.g., an open mouth in screaming or fearful faces of chimpanzees and humans). Therefore, in general, ape face-scanning patterns seem sensitive to variation in faces, whereas humans' face-scanning patterns seem sensitive to particular facial

features; i.e. eyes.

Further comparative studies on face scanning may reveal the similarities and differences between apes and other primate species. For example, the strong tendency of apes and humans to view the internal features of faces is remarkable (although it is somewhat weaker among apes than among humans), and successive fixations on internal features are common in these species (i.e., successive on-feature fixations rather than repetition of on- and off-feature fixations; see Figure 2-6-7). Direct quantitative comparison with macaque species in terms of viewing time and fixation sequences may reveal differences between macaques and apes/humans.

In conclusion, we analyzed typical face-scanning patterns among humans and the three species of great apes. The results indicated that all species are strikingly similar, but humans are somewhat differentiated from the other apes, especially in their eye-viewing patterns. We established the eye-tracking methodology of great apes, which may be useful in future research to further elucidate the similarities and differences in perceptual and cognitive abilities among human and nonhuman primates.

References

Adachi, I., Chou, D. P., & Hampton, R. R. (2009). Thatcher effect in monkeys demonstrates conservation of face perception across primates. *Curr Biol, 19*(15), 1270-1273.

Ankel-Simons, F. (2000). *Primate anatomy: an introduction*. San Diego: Academic Press.

Armann, R., & Bülthoff, I. (2009). Gaze behavior in face comparison: The roles of sex, task, and symmetry. *Attention, Perception, & Psychophysics, 71*(5), 1107-1126.

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PLoS ONE, 3*(8), e3022.

Dahl, C. D., Wallraven, C., Bulthoff, H. H., & Logothetis, N. K. (2009). Humans and macaques employ similar face-processing strategies. *Curr Biol, 19*(6), 509-513.

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nat Neurosci, 8*, 519-526.

Delerue, C., Lapr vote, V., Verfaillie, K., & Boucart, M. (2010). Gaze control during face exploration in schizophrenia. *Neurosci lett, 482*, 245-249.

Ekman, P., & Friesen, W. V. (1978). *The facial action coding system (FACS): A technique for the measurement of facial action.* Palo Alto: Consulting Psychologists Press.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *J Vis, 8*(2), 1-17.

Gothard, K. M., Brooks, K. N., & Peterson, M. A. (2009). Multiple perceptual strategies used by macaque monkeys for face recognition. *Anim Cogn, 12*(1), 155-167.

Green, M. J., Williams, L. M., & Davidson, D. (2003). Visual scanpaths to threat-related faces in deluded schizophrenia. *Psychiat Res, 119*(3), 271-285.

Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y., & Young, M. P. (2003). How do monkeys view faces?—a study of eye movements. *Exp Brain Res, 150*(3), 363-374.

Hainline, L. (1978). Developmental changes in visual scanning of face and nonface patterns by infants. *J Exp Child Psychol, 25*(1), 90-115.

Haith, M. M., Bergman, T., & Moore, M. J. (1977). Eye contact and face scanning in early infancy. *Science, 198*(4319), 853-855.

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *J Vis, 8*(1), 1-10.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Mem Cogn, 33*(1), 98-106.

Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., & Fujita, S. (2010). Facial perception of conspecifics: Chimpanzees (*Pan troglodytes*) preferentially attend to proper orientation and open eyes. *Anim Cogn, 13*(5), 679-688.

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2004). Face to face: visual scanpath evidence for abnormal processing of facial expressions in social phobia. *Psychiat Res, 127*, 43-53.

Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychol Sci, 19*(10), 998-1006.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*, 1-19.

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: a comparative eye-tracking study. *Proc Roy Soc B, 276*(1664), 1949-1955.

Kano, F., & Tomonaga, M. (2010). Face scanning in chimpanzees and humans: Continuity and discontinuity. *Anim Behav, 79*(1), 227-235.

Keating, C. F., & Keating, E. G. (1982). Visual scan patterns of rhesus monkeys viewing faces.

Perception, 11(2), 211-219.

Keating, C. F., & Keating, E. G. (1993). Monkeys and mug shots: Cues used by rhesus monkeys (*Macaca Mulatta*) to recognize a human face. *J Comp Psychol, 107*(2), 131-139.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Arch Gen Psychiatry, 59*(9), 809-816.

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *J Hum Evol, 40*(5), 419-435.

Kuwahata, H., Adachi, I., Fujita, K., Tomonaga, M., & Matsuzawa, T. (2004). Development of schematic face preference in macaque monkeys. *Behav Process, 66*(1), 17-21.

Kuze, N., Malim, T. P., & Kohshima, S. (2005). Developmental changes in the facial morphology of the Borneo Orangutan (*Pongo pygmaeus*): Possible signals in visual communication. *Am J Primatol, 65*(4), 353-376.

Malcolm, G. L., Lanyon, L. J., Fugard, A. J. B., & Barton, J. J. S. (2008). Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects. *J Vis, 8*(8), 1-9.

Matsuzawa, T., Tomonaga, M., & Tanaka, M. (2006). *Cognitive development in chimpanzees*. Tokyo: Springer.

Mendelson, M. J., Haith, M. M., & Goldmanrakic, P. S. (1982). Face scanning and responsiveness to social cues in infant rhesus monkeys. *Dev Psychol, 18*(2), 222-228.

Myowa-Yamakoshi, M., Yamaguchi, M. K., Tomonaga, M., Tanaka, M., & Matsuzawa, T. (2005). Development of face recognition in infant chimpanzees (*Pan troglodytes*). *Cogn Dev, 20*(1), 49-63.

Nahm, F. K. D., Perret, A., Amaral, D. G., & Albright, T. D. (1997). How do monkeys look at faces? *J Cogn Neurosci, 9*(5), 611-623.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *J Autism Dev Disord, 32*(4), 249-261.

Phillips, M. L., & David, A. S. (1997). Visual scan paths are abnormal in deluded schizophrenics. *Neuropsychologia, 35*(1), 99-105.

Stacey, P. C., Walker, S., & Underwood, J. D. M. (2005). Face processing and familiarity: Evidence from eye-movement data. *Br J Psychol, 96*(4), 407-422.

Valenza, E., Simion, F., Cassia, V. M., & Umilt, C. (1996). Face preference at birth. *J Exp Psychol Hum Percept Perform, 22*(4), 892-903.

van der Geest, J. N., Kemner, C., Verbaten, M. N., & van Engeland, H. (2002). Gaze behavior of children with pervasive developmental disorder toward human faces: A fixation time study. *J Child Psychol Psychiat, 43*(5), 669-678.

Tables

Table 2-6-1. Age, sex, and rearing history of each ape.

	Sex	Age (years)	Rearing History
Chimpanzee (KUPRI)			
Ai	F	33	Nursery
Ayumu	M	9	Mother
Chloe	F	29	Nursery
Cleo	F	9	Mother
Pal	F	9	Mother
Pan	F	26	Nursery
Gorilla (WKPRC)			
Gorgo	M	29	Nursery
Kibara	F	6	Mother
Louna	F	4	Mother
Viringika	F	15	Mother
Zola	F	2	Mother
Orangutan (WKPRC)			
Batak	M	1 (+ 1 month)	Mother
Bimbo	M	30	Nursery
Dukana	F	21	Mother
Kila	F	10	Mother
Maia	F	3	Mother
Padana	F	13	Mother
Pini	F	22	Mother
Raja	F	7	Mother
Suaq	M	1 (+ 6 months)	Mother
Tanah	F	1 (+ 5 months)	Mother

Table 2-6-2. Proportion of viewing time for each part of a scene (not normalized for size) in each species across studies.

	Conspecific scene			**Allospecific scene**		
Participant (Study)	Face	Body	Background	Face	Body	Background
Human Adult At WKPRC (1)	0.7	0.21	0.08	0.66	0.28	0.05
Human Adult At KUPRI (2)	0.57	0.21	0.21	0.72	0.18	0.09
Chimpanzee (2)	0.37	0.43	0.18	0.31	0.34	0.34
Gorilla (1)	0.42	0.38	0.25	0.33	0.34	0.32
Orangutan (1)	0.45	0.29	0.19	0.34	0.43	0.22

(1) This study

(2) Kano and Tomonaga (2010)

Table 2-6-3. Proportion of viewing time for each part of a face (not normalized for size) in each species across studies.

	Conspecific face				**Allospecific face**			
Participant (Study)	Eye (Eyeball)	Nose	Mouth	Periphery	Eye (Eyeball)	Nose	Mouth	Periphery
Human Adult At WKPRC (1)	0.64 (0.23)	0.23	0.04	0.07	0.68 (0.26)	0.15	0.08	0.07
Human Adult At KUPRI (2)	0.44 (0.27)	0.17	0.19	0.18	0.51 (0.28)	0.18	0.13	0.16
Chimpanzee (2)	0.37 (0.11)	0.16	0.25	0.2	0.28 (0.07)	0.24	0.15	0.31
Gorilla (1)	0.48 (0.18)	0.2	0.13	0.17	0.41 (0.13)	0.07	0.18	0.31
Orangutan (1)	0.36 (0.15)	0.13	0.19	0.29	0.38 (0.12)	0.12	0.2	0.29

(1) This study

(2) Kano and Tomonaga (2010)

Table 2-6-4. Edit costs of two given scanpaths (four-letter length) between trials and individuals. Italics in ape participants' cells indicate significant differences from human participants ($P < 0.05$).

	Participant						
		WKPRC		**KUPRI**		**all**	
Stimuli	Human	Gorilla	Orangutan	Human	Chimpanzee	Human	Ape
Between trials (within individual)	2.23	2.40	2.46	1.84	*2.13*	2.03	*2.40*
Between individuals (within species)	2.13	*2.29*	*2.47*	1.97	*2.14*	2.05	*2.45*

Figure captions

Figure 2-6-1. Apes and apparatus. (a) Eye-tracking apparatus. An eye tracker and a monitor are mounted on the movable platform (front). The experimenter controlled the apparatus using the computer mounted on the other platform (behind). See also: http://www.youtube.com/watch?v=zHx2KwQEHq0. (b) A juvenile gorilla drinking grape juice via a tube attached to the transparent acrylic panel. (c) An adult female orangutan on the apparatus.

Figure 2-6-2. Examples of stimuli: (a) whole-body pictures; (b) facial pictures; (c) example areas of interest (AOI) for facial pictures.

Figure 2-6-3. Scanning pattern for whole-body pictures by each species. (a) Proportion of viewing time for each AOI with respect to the total scene viewing time. (b) Probability of fixation for each AOI across fixation order. All data are normalized for area sizes. Error bars represent the upper and lower bounds of the 95% confidence intervals.

Figure 2-6-4. Scanning pattern for facial pictures by each species. (a) Proportion of viewing time for each AOI with respect to total face-viewing time. (b) Probability of fixation for each AOI across fixation order. All data are normalized for area sizes. Pictures of juvenile gorillas and orangutans are not included in these data. Error bars represent the upper and lower bounds of the 95% confidence intervals.

Figure 2-6-5. The proportion of viewing time for each AOI with respect to total face-viewing time as a function of face types in gorillas and orangutans. All data are normalized for area sizes. Error bars represent the upper and lower bounds of the 95% confidence intervals. Asterisk indicates $P < 0.05$ in post hoc tests.

Figure 2-6-6. Typical scanning pattern for facial pictures. Probabilities of saccade destination are presented in percentages and scaled to size. If a WKPRC human is currently looking at the right eye of a human face (top center), he/she re-fixates the right eye with a probability of 2.6, fixates the left

eye with probability of 15.6, fixates the nose with probability of 7.7, etc. (see text for details). No direction is given for connections between AOIs. Pictures of juvenile gorillas and orangutans are not included in these data.

Figure 2-6-7. Example scanning patterns for facial pictures. The eye-movement paths of five participants from each species were superimposed on the facial pictures.

Figures

Figure 2-6-1

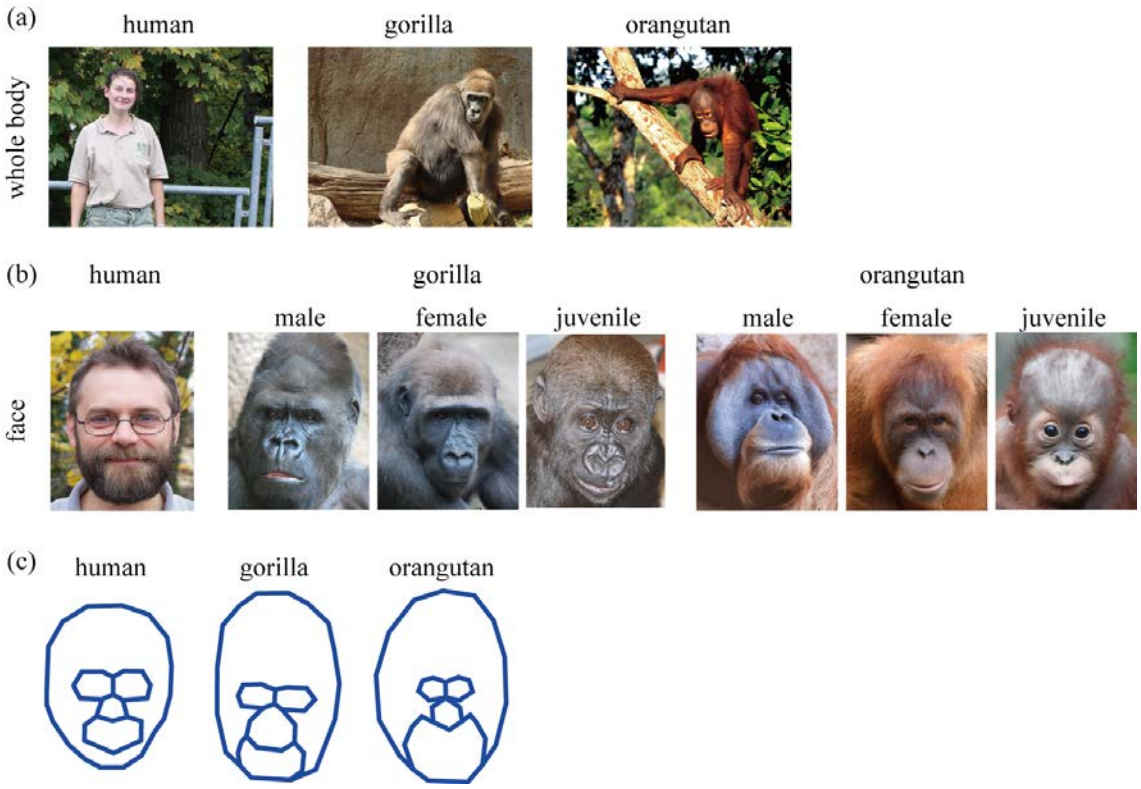

(a)
human
gorilla
orangutan
whole body
(b)
human
gorilla
orangutan
male
female
juvenile
male
female
juvenile
face
(c)
human
gorilla
orangutan

Figure 2-6-2

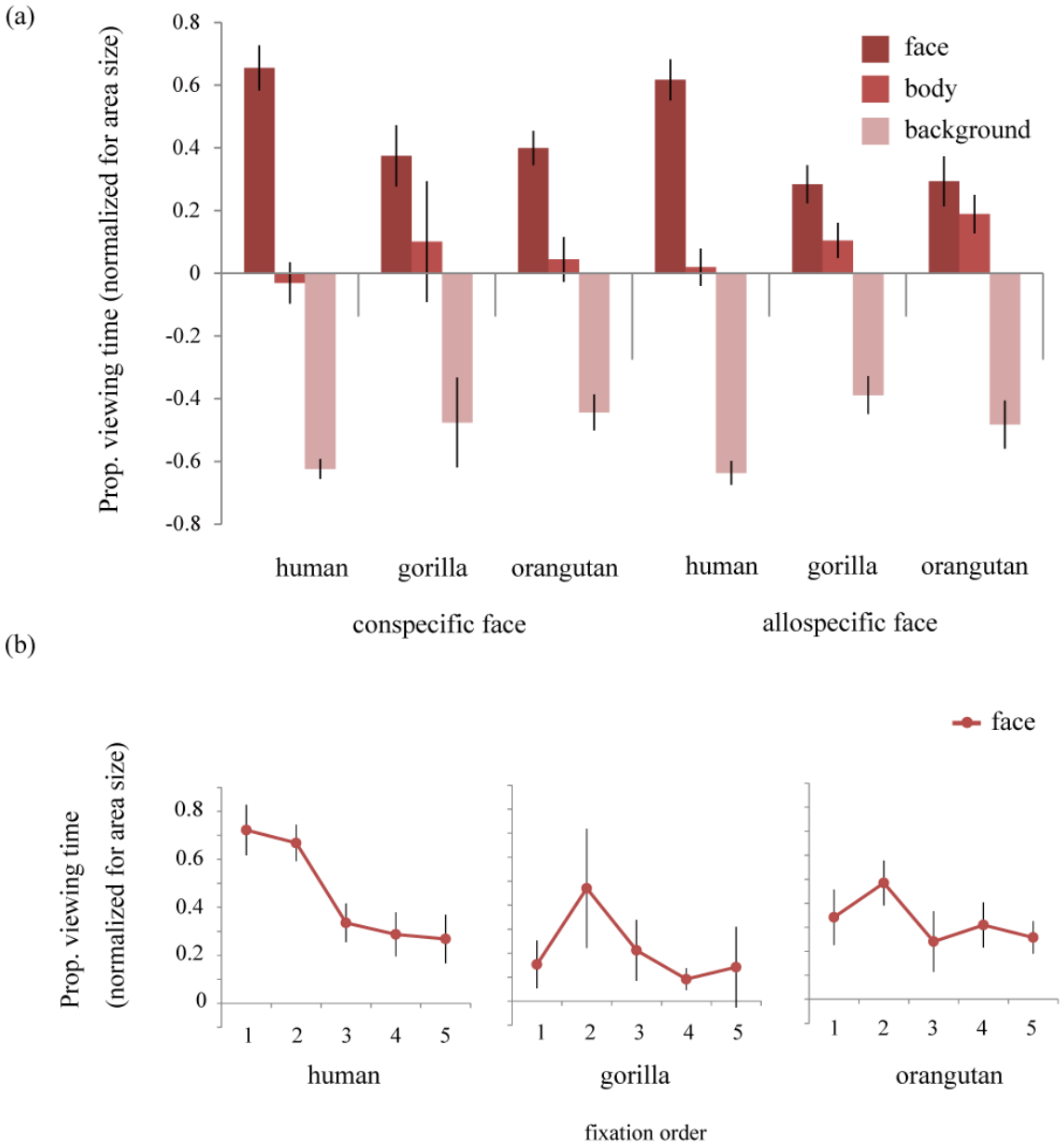

(a)
Prop. viewing time (normalized for area size)
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
face
body
background
human
gorilla
orangutan
human
gorilla
orangutan
conspecific face
allospecific face
(b)
Prop. viewing time
(normalized for area size)
face
0.8
0.6
0.4
0.2
0
1
2
3
4
5
human
gorilla
orangutan
fixation order

Figure 2-6-3

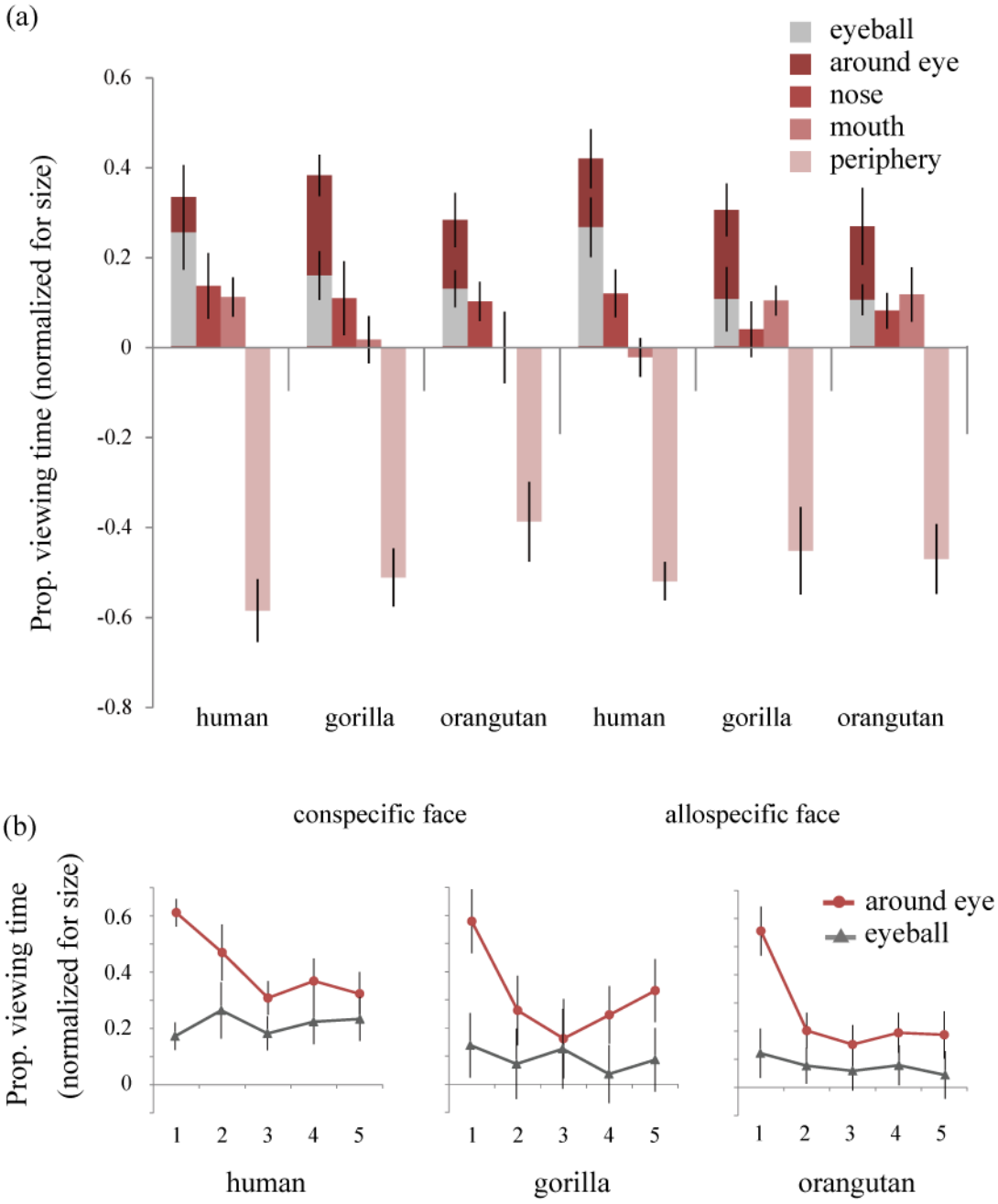

(a)
eyeball
around eye
nose
mouth
periphery
Prop. viewing time (normalized for size)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
human
gorilla
orangutan
human
gorilla
orangutan
(b)
conspecific face
allospecific face
Prop. viewing time (normalized for size)
around eye
eyeball
0.6
0.4
0.2
0
1 2 3 4 5
human
gorilla
orangutan

Figure 2-6-4

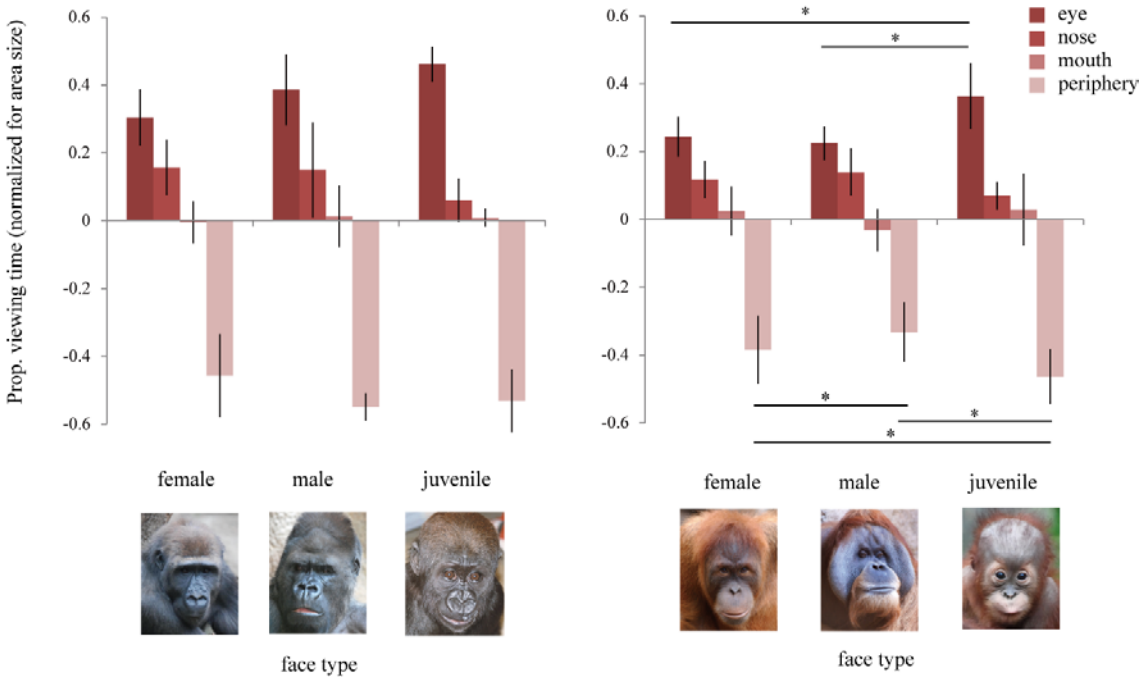

Prop. viewing time (normalized for area size)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
female
male
juvenile
face type
eye
nose
mouth
periphery
*
*
*
*
*
female
male
juvenile
face type

Figure 2-6-5

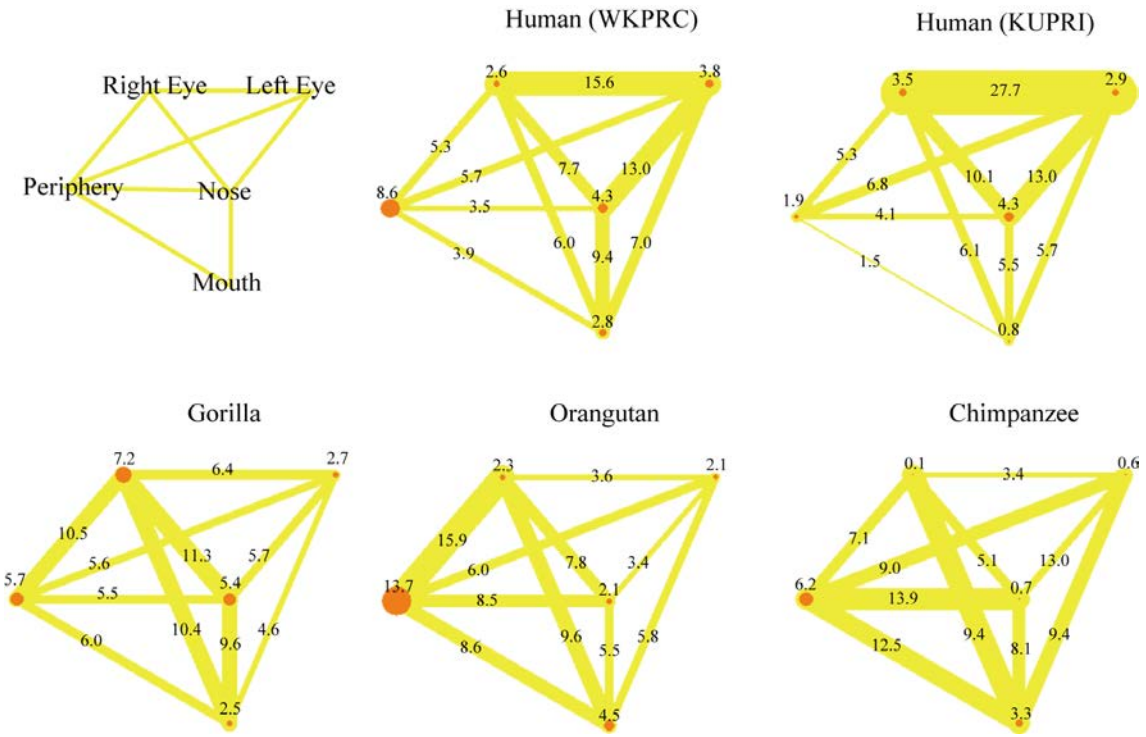

Right Eye
Left Eye
Periphery
Nose
Mouth
Human (WKPRC)
2.6
15.6
3.8
5.3
5.7
7.7
13.0
8.6
4.3
3.5
6.0
7.0
3.9
9.4
2.8
Human (KUPRI)
3.5
27.7
2.9
5.3
6.8
10.1
13.0
1.9
4.3
4.1
6.1
5.7
1.5
5.5
0.8
Gorilla
7.2
6.4
2.7
10.5
5.6
11.3
5.7
5.7
5.4
5.5
10.4
9.6
4.6
6.0
2.5
Orangutan
2.3
3.6
2.1
15.9
6.0
7.8
3.4
13.7
8.5
2.1
9.6
5.8
8.6
5.5
4.5
Chimpanzee
0.1
3.4
0.6
7.1
9.0
5.1
13.0
6.2
13.9
0.7
9.4
9.4
12.5
8.1
3.3

Figure 2-6-6

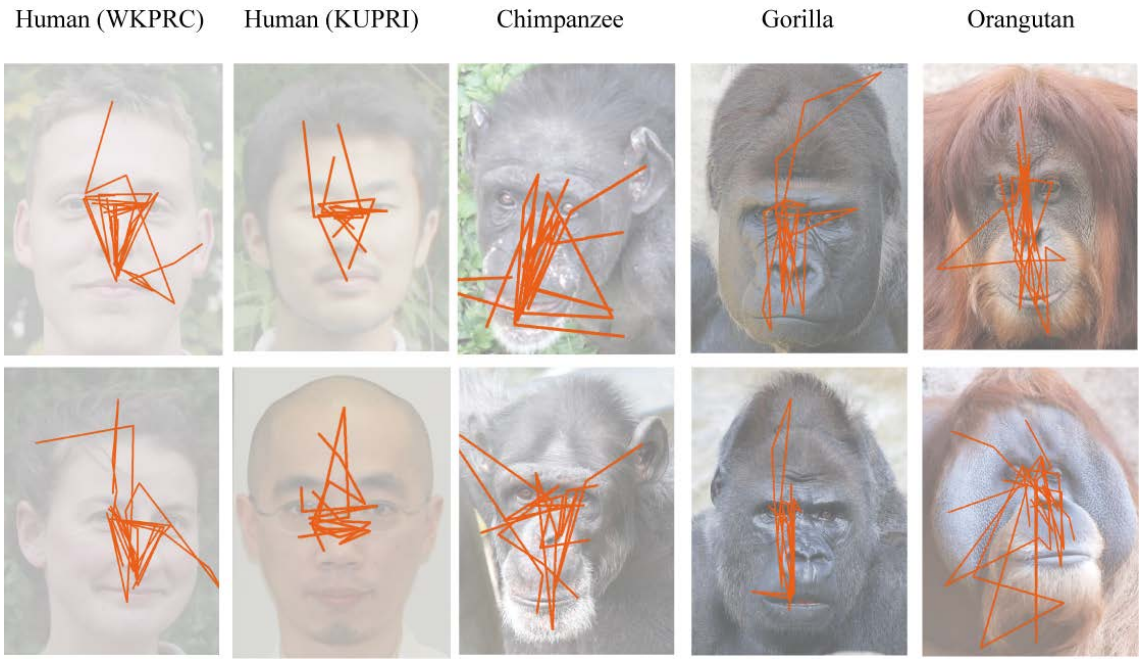


Figure 2-6-7

Acknowledgement

These studies were financially supported by the Japan Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan Grants-in-Aid for Scientific Research (nos. 16002001, 19300091, 20002001, 20220004, 20680015, 212299, 23300103), the JSPS/MEXT global COE programs (D07 and A06), and the JSPS Institutional Program for Young Researcher Overseas Visits. I thank my supervisors, Drs. M. Tomonaga and T. Matsuzawa for their generous support and encouragement. I also thank M. Tanaka, M. Hayashi, I. Adachi, S. Hirata, C. Hashimoto, J. Call, Y. Hattori, T. Kaneko, and C. Martin for their kind collaborations and fruitful discussions. Special thanks are due to C. Rahn, F. Stock, H. Knofe, J. Corbit, R. Pieszek, and the other members of Wolfgang Köhler Primate Research Center; K. Fuwa, K. Sugama, K. Kusunoki, and S. Fujita the other members of Great Ape Research Institute, Hayashibara Biomedical Laboratories. Inc.; T. Tomoko, M. Nakashima, M. Sakai, Y. Okumura, and the other members of Primate Research Institute for their help with the experiment. Thanks are also due to the Leipzig Zoo in Germany, the Hayashibara Biomedical Laboratory, Inc. in Japan, and the Center for Human Evolution Modeling Research at the Primate Research Institute in Japan for the daily care of the apes. Lastly, I deeply thank the reviewers of this thesis for their helpful comments.